DANIEL P. LOUCKS and
EELCO VAN BEEK

# WATER RESOURCES SYSTEMS PLANNING AND MANAGEMENT

## An Introduction to Methods, Models and Applications

with contributions from
Jery R. Stedinger
Jos P.M. Dijkman
Monique T. Villars

United Nations
Educational, Scientific and
Cultural Organization

Studies and Reports in Hydrology

Water Resources Systems Planning
and Management

An Introduction to Methods,
Models and Applications

# Water Resources Systems Planning and Management

## An Introduction to Methods, Models and Applications

Daniel P. Loucks and Eelco van Beek
with contributions from

Jery R. Stedinger
Jozef P.M. Dijkman
Monique T. Villars

Studies and Reports in Hydrology

UNESCO PUBLISHING

Published in 2005 by the United Nations Educational,
Scientific and Cultural Organization
7, place de Fontenoy F-75352 Paris 07 SP

Typeset by SR Nova Pvt. Ltd, Bangalore, India

Printed by Ages Arti Grafiche, Turin

ISBN 92-3-103998-9


*Printed in Italy*

# Foreword

Within the Netherlands, as in much of the world, the quality of our lives is directly related to the quality of our natural environment – our air, land and water resources. We consider a quality environment crucial to human health and economic and social development as well as for ecosystem preservation and diversity. How well we manage our natural resources today will determine just how well these resources will serve our descendants and us. Hence, we care much about the management of these resources, especially our water resources.

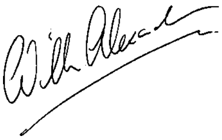
Many of us in the Netherlands are living in areas that exist only because of the successful efforts of our past water engineers, planners and managers. Managing water in ways that best meet all our diverse needs for water and its protection, including the needs of natural ecosystems, is absolutely essential. But in spite of our knowledge and experience, we Dutch, as others throughout the world, still experience droughts, floods and water pollution. These adverse impacts are not unique to us here in Europe. In too many other regions of this world the need for improved water management is much more critical and much more urgent. Too many people, especially children, suffer each day because of the lack of it.

As we take pride in our abilities to manage water, we also take pride in our abilities to help others manage water. Institutions such as WL | Delft Hydraulics have been doing this throughout its seventy-five years of existence. This book was written and published, in part, to celebrate its seventy-fifth anniversary.

This book was written by individuals who have simultaneously served as university professors as well as consulting engineers throughout much of their professional careers. They have provided an introduction to practical ways of modeling and analysing water resources systems.

Whether you are studying at a university or working in a developed or developing region, the methods and advice presented in this book can help you develop your skills in the use of quantitative methods of identifying and evaluating effective water resources management plans and policies. It can serve as a guide on ways of obtaining the information you and your organization need when deciding how to best manage these important resources.

This book, introducing an integrated systems approach to water management, can serve many students, teachers, and practising water resource engineers and planners in the years to come.



**His Royal Highness the Prince of Orange**
**The Netherlands**

# Preface

Throughout history much of the world has witnessed ever-greater demands for reliable, high-quality and inexpensive water supplies for domestic consumption, agriculture and industry. In recent decades there have also been increasing demands for hydrological regimes that support healthy and diverse ecosystems, provide for water-based recreational activities, reduce if not prevent floods and droughts, and in some cases, provide for the production of hydropower and ensure water levels adequate for ship navigation. Water managers are challenged to meet these multiple and often conflicting demands. At the same time, public stakeholder interest groups have shown an increasing desire to take part in the water resources development and management decision-making process. Added to all these management challenges are the uncertainties of natural water supplies and demands due to changes in our climate, changes in people's standards of living, changes in watershed land uses and changes in technology. How can managers develop, or redevelop and restore, and then manage water resources systems – systems ranging from small watersheds to those encompassing large river basins and coastal zones – in a way that meets society's changing objectives and goals? In other words, how can water resources systems become more integrated and sustainable?

Before engineering projects can be undertaken to address water management problems or to take advantage of opportunities for increased economic, ecological, environmental and social benefits, they must first be planned. This involves identifying various alternatives for addressing the problems or opportunities. Next, the various impacts of each proposed alternative need to be estimated and evaluated. A variety of optimization and simulation models and modelling approaches have been developed to assist water planners and managers in identifying and evaluating plans. This book introduces the science and art of modelling in support of water resources planning and management. Its main emphasis is on the practice of developing and using models to address specific water resources planning and management problems. This must be done in ways that provide relevant, objective and meaningful information to those who are responsible for making informed decisions about specific issues in specific watersheds or river basins.

Readers of this book are not likely to learn this art of modelling unless they actually employ it. The information, examples and case studies contained in this book, together with the accompanying exercises, we believe, will facilitate the process of becoming a skilled water resources systems modeller, analyst and planner. This has been our profession, and we can highly recommend it to others. Planning and management modelling is a multidisciplinary activity that is an essential part of almost all projects designed to increase the benefits, however measured, from available water and related land resources. The modelling and analysis of water resources systems involves science and it also involves people and politics. It is a challenge, but it is also fulfilling.

This book builds on a text titled *Water Resources Systems Planning and Analysis* by Loucks, Stedinger and Haith published by Prentice Hall in 1981. The present work updates much of what was in that text, introduces some new modelling methods that are proving to be useful, and contains considerably more case studies. It benefits considerably from the experiences of WL | Delft Hydraulics, one of the many firms involved around the world using the approaches and methods discussed in this book.

Developments in graphics-based menu-driven interactive computer programs and computer technology during the last quarter of a century have had a significant and beneficial impact on the use of modelling in the practice of water resources engineering, planning and management. All the models discussed in this book are designed for use on micro-computers. The software we use to illustrate the solutions to various problems can be obtained from the Internet free of charge. Commonly available spreadsheet software can also be used. None of this was available in 1981.

Although we have attempted to incorporate into each chapter current approaches to water resources systems planning and analysis, this book does not pretend to be a review of the state-of-the-art of water resources systems analysis found in the literature. Rather it is intended to introduce readers to some of the more commonly used models and modelling approaches applied to the planning and managing of water resources systems. We have tried to organize our discussion of these topics in a way useful for teaching and self-study. The contents reflect our belief that the most appropriate methods for planning and management are often the simpler ones, chiefly because they are easier to understand and explain, require less input data and time, and are easier to apply to specific issues or problems. This does not imply that more sophisticated and complex models are less useful. Sometimes their use is the only way one can provide the needed information. In this book, we attempt to give readers the knowledge to make appropriate choices regarding model complexity. These choices will depend in part on factors such as the issues being addressed and the information needed, the level of accuracy desired, the availability of data and their cost, and the time required and available to carry out the analysis. While many analysts have their favourite modelling approach, the choice of model should be based on a knowledge of various modelling approaches and their advantages and limitations.

This book assumes readers have had some mathematical training in algebra, calculus, geometry and the use of vectors and matrices. Readers of Chapters 7 through 9 will benefit from some background in probability and statistics. Similarly, some exposure to micro-economic theory and welfare economics will be useful for readers of Chapter 10. Some knowledge of hydrology, hydraulics and environmental engineering will also be beneficial, but not absolutely essential. Readers wanting an overview of some of natural processes that take place in watersheds, river basins, estuaries and coastal zones can refer to Appendix A. An introductory course in optimization and simulation methods, typically provided in either an operations research or an economic theory course, can also benefit the reader, but again it is not essential.

Chapter 1 introduces water resources systems planning and management and describes some examples of water resources systems projects in which modelling has had a critical role. These example projects also serve to identify some of the current issues facing water managers in different parts of the world. Chapter 2 defines the modelling approach in general and the role of models in water resources planning and management projects. Chapter 3 begins the discussion of optimization and simulation modelling methods and how they are applied and used in practice. It also discusses how modelling activities in water resources development, planning and/or management projects should be managed.

Chapter 4 is devoted to optimization modelling. This relatively large chapter focuses on the use of various optimization methods for the preliminary definition of infrastructure design and operating policies. These preliminary results define alternatives that usually need to be further analysed and improved using simulation methods. The advantages and limitations of different optimization approaches are presented and illustrated using some simple water allocation, reservoir operation and water quality management problems. Chapter 5 extends this discussion of optimization to problems characterized by 'fuzzy' (more qualitative) objectives.

Chapter 6 introduces some of the more recently developed methods of statistical modelling, including artificial neural networks and evolutionary search methods including genetic algorithms. This chapter expects interested readers will refer to other books, many of which are solely devoted to just these topics, for more detail.

Chapters 7 through 9 are devoted to probabilistic models, uncertainty and sensitivity analyses. These methods are useful not only for identifying more realistic infrastructure designs and operating policies given hydrological variability and uncertain parameter values and objectives, but also for estimating some of the major uncertainties associated with model predictions. Such probabilistic and stochastic models can also help identify just what model input data are needed and how accurate those data need be with respect to their influence on the decisions being considered.

Water resources planning and management today inevitably involve multiple goals or objectives, many of which may be conflicting. It is difficult, if not impossible, to please all stakeholders all the time. Models containing multiple objectives can be used to identify the tradeoffs among conflicting objectives. This is information useful to decision-makers who must decide what the best tradeoffs should be, both among conflicting objectives and among conflicting stakeholder interest groups. Multi-objective modelling, Chapter 10, identifies various types of economic, environmental and physical objectives, and some commonly used ways of including multiple objectives in optimization and simulation models.

Chapter 11 is devoted to various approaches for modelling the hydrological processes in river basins. The focus is on water quantity prediction and management. This is followed by Chapter 12 on the prediction and management of water quality processes in river basins and Chapter 13 on the prediction and management of water quantity and quality in storm water runoff, water supply distribution and treatment, and wastewater collection and treatment in urban areas. The final Chapter (Chapter 14) provides a synopsis, reviewing again the main role of models, introducing measures that can be used to evaluate their usefulness in particular projects, and presenting some more case studies showing the application of models to water resources management issues and problems.

Following these fourteen chapters are five appendices. They contain descriptions of A) natural hydrological and ecological processes in river basins, estuaries and coastal zones, B) monitoring and adaptive management, C) drought management, D) flood management, and E) a framework for assessing, developing and managing water resources systems as practiced by WL | Delft Hydraulics.

We believe Chapters 1 through 4 are useful prerequisites to most of the remaining chapters. For university teachers, the contents of this book represent more than can normally be covered in a single quarter or semester course. A first course can include Chapters 1 through 4, and possibly Chapters 10, 11 or 13 in addition to Chapter 14, depending on the background of the participants in the class. A second course could include Chapters 7 through 9 and/or any combination of Chapters 5, 6, 12, 13 or 14, as desired. Clearly much depends on the course objectives and on the background knowledge of the course participants. Some exercises for each chapter are included in the attached CD. (Instructors may write to the authors to obtain suggested solutions to these exercises.)

The writing of this book began at WL | Delft Hydraulics as a contribution to its seventy-fifth anniversary. We are most grateful for the company's support, both financial and intellectual. While this book is not intended to be a testimony to Delft Hydraulics' contributions to the development and application of models to water resources planning and management projects, it does reflect the approaches taken, and modelling tools used by them and other such firms and organizations that engage in water resources planning, development and management projects worldwide.

Many have helped us prepare this book. Jery Stedinger wrote much of Chapters 7, 8 and 9, Nicki Villars helped substantially with Chapter 12, and Jozef Dijkman contributed a major portion of Appendix D. Vladam Babovic, Henk van den Boogaard, Tony Minns, and Arthur Mynett contributed material for Chapter 6. Roland Price provided material for Chapter 13. Others who offered advice and who helped review earlier chapter drafts include Martin Baptist, Herman Breusers, Harm Duel, Herman Gerritsen, Peter Gijsbers, Jos van Gils, Simon Groot, Karel Heynert, Joost Icke, Hans Los, Marcel Marchand, Erik Mosselman, Erik Ruijgh, Johannes Smits, Mindert de Vries and Micha Werner. Ruud Ridderhof and Engelbert Vennix created the figures and tables in this book. We thank all these individuals and others, including our students, who provided assistance and support on various aspects during the entire time this book was being prepared. We have also benefited from the comments of Professors Jan-Tai Kuo at National Taiwan University in Taipei, Jay Lund at the University of California at Davis, Daene McKinney of the University of Texas in Austin, Peter Rogers at Harvard University in Cambridge, Massachusetts, Tineke Ruijgh at TU-Delft, and Robert Traver at Villanova University in Philadelphia, all of whom have used earlier drafts of this book in their classes. Finally we acknowledge with thanks the support of Andras Szöllösi-Nagy and the publishing staff at UNESCO for publishing and distributing this book as a part of their International Hydrological Programme. We have written this book for an international audience, and hence we are especially grateful for, and pleased to have, this connection to and support from UNESCO.

Most importantly we wish to acknowledge and thank all our teachers, students and colleagues throughout the world who have taught us all we know and added to the quality of our professional and personal lives. We have tried our best to make this book error free, but inevitably somewhere there will be flaws. For that we apologize and take responsibility for any errors of fact, judgment or science that may be contained in this book. We will be most grateful if you let us know of any or have other suggestions for improving this book.

***Daniel. P. Loucks,***
Cornell University, Ithaca, N.Y., USA

***Eelco van Beek,***
WL | Delft Hydraulics, Delft,
the Netherlands
November 2004

# Contents

Introduction xix

1. Water Resources Planning and Management: An Overview 3

1. Introduction 3
2. Planning and Management Issues: Some Case Studies 4
   - 2.1. Kurds Seek Land, Turks Want Water 4
   - 2.2. Sharing the Water of the Jordan River Basin: Is There a Way? 6
   - 2.3. Mending the ‘Mighty and Muddy’ Missouri 7
   - 2.4. The Endangered Salmon 7
   - 2.5. The Yellow River: How to Keep the Water Flowing 9
   - 2.6. Lake Source Cooling: Aid to Environment or Threat to Lake? 10
   - 2.7. Managing Water in the Florida Everglades 11
   - 2.8. Restoration of Europe’s Rivers and Seas 13
     - 2.8.1. The Rhine 13
     - 2.8.2. The Danube 14
     - 2.8.3. The North and Baltic Seas 15
   - 2.9. Egypt and the Nile: Limits to Agricultural Growth 16
   - 2.10. Damming the Mekong 16
3. So, Why Plan, Why Manage? 18
   - 3.1. Too Little Water 20
   - 3.2. Too Much Water 20
   - 3.3. Polluted Water 21
   - 3.4. Degradation of Aquatic and Riparian Ecosystems 21
   - 3.5. Other Planning and Management Issues 21
4. System Components, Planning Scales and Sustainability 22
   - 4.1. Spatial Scales for Planning and Management 22
   - 4.2. Temporal Scales for Planning and Management 23
   - 4.3. Sustainability 23
5. Planning and Management 24
   - 5.1. Approaches 24
     - 5.1.1. Top-Down Planning and Management 25
     - 5.1.2. Bottom-Up Planning and Management 25
     - 5.1.3. Integrated Water Resources Management 26
   - 5.2. Planning and Management Aspects 26
     - 5.2.1. Technical Aspects 26
     - 5.2.2. Economic and Financial Aspects 27
     - 5.2.3. Institutional Aspects 28
   - 5.3. Analyses for Planning and Management 28
   - 5.4. Models for Impact Prediction and Evaluation 30
   - 5.5. Shared-Vision Modelling 31
   - 5.6. Adaptive Integrated Policies 31
   - 5.7. Post-Planning and Management Issues 32

6. Meeting the Planning and Management Challenges: A Summary 32
7. References 34

2. Water Resource Systems Modelling: Its Role in Planning and Management 39
1. Introduction 39
2. Modelling of Water Resources Systems 41
2.1. An Example Modelling Approach 41
2.2. Characteristics of Problems to be Modelled 41
3. Challenges in Water Resources Systems Modelling 43
3.1. Challenges of Planners and Managers 43
3.2. Challenges of Modelling 44
3.3. Challenges of Applying Models in Practice 45
4. Developments in Modelling 46
4.1. Modelling Technology 46
4.2. Decision Support Systems 47
4.2.1. Shared-Vision Modelling 49
4.2.2. Open Modelling Systems 51
4.2.3. Example of a DSS for River Flood Management 51
5. Conclusions 54
6. References 55

3. Modelling Methods for Evaluating Alternatives 59
1. Introduction 59
1.2. Model Components 60
2. Plan Formulation and Selection 61
2.1. Plan Formulation 61
2.2. Plan Selection 63
3. Modelling Methods: Simulation or Optimization 64
3.1. A Simple Planning Example 65
3.2. Simulation Modelling Approach 66
3.3. Optimization Modelling Approach 66
3.4. Simulation Versus Optimization 67
3.5. Types of Models 69
3.5.1. Types of Simulation Models 69
3.5.2. Types of Optimization Models 70
4. Model Development 71
5. Managing Modelling Projects 72
5.1. Creating a Model Journal 72
5.2. Initiating the Modelling Project 72
5.3. Selecting the Model 73
5.4. Analysing the Model 74
5.5. Using the Model 74
5.6. Interpreting Model Results 75
5.7. Reporting Model Results 75
6. Issues of Scale 75
6.1. Process Scale 75
6.2. Information Scale 76
6.3. Model Scale 76
6.4. Sampling Scale 76
6.5. Selecting the Right Scales 76
7. Conclusions 77
8. References 77

4. Optimization Methods 81
1. Introduction 81
2. Comparing Time Streams of Economic Benefits and Costs 81
2.1. Interest Rates 82
2.2. Equivalent Present Value 82
2.3. Equivalent Annual Value 82
3. Non-linear Optimization Models and Solution Procedures 83
3.1. Solution Using Calculus 84
3.2. Solution Using Hill Climbing 84
3.3. Solution Using Lagrange Multipliers 86
3.3.1. Approach 86
3.3.2. Meaning of the Lagrange Multiplier 88
4. Dynamic Programming 90
4.1. Dynamic Programming Networks and Recursive Equations 90
4.2. Backward-Moving Solution Procedure 92
4.3. Forward-Moving Solution Procedure 95
4.4. Numerical Solutions 96
4.5. Dimensionality 97
4.6. Principle of Optimality 97
4.7. Additional Applications 97
4.7.1. Capacity Expansion 98
4.7.2. Reservoir Operation 102
4.8. General Comments on Dynamic Programming 112
5. Linear Programming 113
5.1. Reservoir Storage Capacity–Yield Models 114
5.2. A Water Quality Management Problem 117
5.2.1. Model Calibration 118
5.2.2. Management Model 119
5.3. A Groundwater Supply Example 124
5.3.1. A Simplified Model 125
5.3.2. A More Detailed Model 126

5.3.3. An Extended Model 127
5.3.4. Piecewise Linearization Methods 128
5.4. A Review of Linearization Methods 129
6. A Brief Review 132
7. References 132

5. Fuzzy Optimization 135
1. Fuzziness: An Introduction 135
1.1. Fuzzy Membership Functions 135
1.2. Membership Function Operations 136
2. Optimization in Fuzzy Environments 136
3. Fuzzy Sets for Water Allocation 138
4. Fuzzy Sets for Reservoir Storage and Release Targets 139
5. Fuzzy Sets for Water Quality Management 140
6. Summary 144
7. Additional References (Further Reading) 144

6. Data-Based Models 147
1. Introduction 147
2. Artificial Neural Networks 148
2.1. The Approach 148
2.2. An Example 151
2.3. Recurrent Neural Networks for the Modelling of Dynamic Hydrological Systems 153
2.4. Some Applications 153
2.4.1. RNN Emulation of a Sewerage System in the Netherlands 154
2.4.2. Water Balance in Lake IJsselmeer 155
3. Genetic Algorithms 156
3.1. The Approach 156
3.2. Example Iterations 158
4. Genetic Programming 159
5. Data Mining 163
5.1. Data Mining Methods 163
6. Conclusions 164
7. References 165

7. Concepts in Probability, Statistics and Stochastic Modelling 169
1. Introduction 169
2. Probability Concepts and Methods 170
2.1. Random Variables and Distributions 170
2.2. Expectation 173
2.3. Quantiles, Moments and Their Estimators 173
2.4. L-Moments and Their Estimators 176
3. Distributions of Random Events 179
3.1. Parameter Estimation 179
3.2. Model Adequacy 182
3.3. Normal and Lognormal Distributions 186
3.4. Gamma Distributions 187
3.5. Log-Pearson Type 3 Distribution 189
3.6. Gumbel and GEV Distributions 190
3.7. L-Moment Diagrams 192
4. Analysis of Censored Data 193
5. Regionalization and Index-Flood Method 195
6. Partial Duration Series 196
7. Stochastic Processes and Time Series 197
7.1. Describing Stochastic Processes 198
7.2. Markov Processes and Markov Chains 198
7.3. Properties of Time-Series Statistics 201
8. Synthetic Streamflow Generation 203
8.1. Introduction 203
8.2. Streamflow Generation Models 205
8.3. A Simple Autoregressive Model 206
8.4. Reproducing the Marginal Distribution 208
8.5. Multivariate Models 209
8.6. Multi-Season, Multi-Site Models 211
8.6.1. Disaggregation Models 211
8.6.2. Aggregation Models 213
9. Stochastic Simulation 214
9.1. Generating Random Variables 214
9.2. River Basin Simulation. 215
9.3. The Simulation Model 216
9.4. Simulation of the Basin 216
9.5. Interpreting Simulation Output 217
10. Conclusions 223
11. References 223

8. Modelling Uncertainty 231
1. Introduction 231
2. Generating Values From Known Probability Distributions 231
3. Monte Carlo Simulation 233
4. Chance Constrained Models 235
5. Markov Processes and Transition Probabilities 236
6. Stochastic Optimization 239
6.1. Probabilities of Decisions 243
6.2. A Numerical Example 244
7. Conclusions 251
8. References 251

9. Model Sensitivity and Uncertainty Analysis 255
1. Introduction 255
2. Issues, Concerns and Terminology 256
3. Variability and Uncertainty In Model Output 258
3.1. Natural Variability 259
3.2. Knowledge Uncertainty 260
3.2.1. Parameter Value Uncertainty 260
3.2.2. Model Structural and Computational Errors 260
3.3. Decision Uncertainty 260
4. Sensitivity and Uncertainty Analyses 261
4.1. Uncertainty Analyses 261
4.1.1. Model and Model Parameter Uncertainties 262
4.1.2. What Uncertainty Analysis Can Provide 265
4.2. Sensitivity Analyses 265
4.2.1. Sensitivity Coefficients 267
4.2.2. A Simple Deterministic Sensitivity Analysis Procedure 267
4.2.3. Multiple Errors and Interactions 269
4.2.4. First-Order Sensitivity Analysis 270
4.2.5. Fractional Factorial Design Method 272
4.2.6. Monte Carlo Sampling Methods 273
5. Performance Indicator Uncertainties 278
5.1. Performance Measure Target Uncertainty 278
5.2. Distinguishing Differences Between Performance Indicator Distributions 281
6. Communicating Model Output Uncertainty 283
7. Conclusions 285
8. References 287

10. Performance Criteria 293
1. Introduction 293
2. Informed Decision-Making 294
3. Performance Criteria and General Alternatives 295
3.1. Constraints On Decisions 296
3.2. Tradeoffs 296
4. Quantifying Performance Criteria 297
4.1. Economic Criteria 298
4.1.1. Benefit and Cost Estimation 299
4.1.2. A Note Concerning Costs 302
4.1.3. Long and Short-Run Benefit Functions 303
4.2. Environmental Criteria 305
4.3. Ecological Criteria 306
4.4. Social Criteria 308
5. Multi-Criteria Analyses 309
5.1. Dominance 310
5.2. The Weighting Method 311
5.3. The Constraint Method 312
5.4. Satisficing 313
5.5. Lexicography 313
5.6. Indifference Analysis 313
5.7. Goal Attainment 314
5.8. Goal-Programming 315
5.9. Interactive Methods 315
5.10. Plan Simulation and Evaluation 316
6. Statistical Summaries of Performance Criteria 320
6.1. Reliability 321
6.2. Resilience 321
6.3. Vulnerability 321
7. Conclusions 321
8. References 322

11. River Basin Planning Models 325
1. Introduction 325
1.1. Scales of River Basin Processes 326
1.2. Model Time Periods 327
1.3. Modelling Approaches for River Basin Management 328
2. Modelling the Natural Resources System and Related Infrastructure 328
2.1. Watershed Hydrological Models 328
2.1.1. Classification of Hydrological Models 329
2.1.2. Hydrological Processes: Surface Water 329
2.1.3. Hydrological Processes: Groundwater 333
2.1.4. Modelling Groundwater: Surface Water Interactions 336
2.1.5. Streamflow Estimation 339
2.1.6. Streamflow Routing 341
2.2. Lakes and Reservoirs 342
2.2.1. Estimating Active Storage Capacity 343
2.2.2. Reservoir Storage–Yield Functions 344
2.2.3. Evaporation Losses 346
2.2.4. Over and Within-Year Reservoir Storage and Yields 347
2.2.5. Estimation of Active Reservoir Storage Capacities for Specified Yields 348
2.3. Wetlands and Swamps 354
2.4. Water Quality and Ecology 354

3. Modelling the Socio-Economic Functions In a River Basin 355
3.1. Withdrawals and Diversions 355
3.2. Domestic, Municipal and Industrial Water Demand 356
3.3. Agricultural Water Demand 357
3.4. Hydroelectric Power Production 357
3.5. Flood Risk Reduction 359
3.5.1. Reservoir Flood Storage Capacity 360
3.5.2. Channel Capacity 362
3.6. Lake-Based Recreation 362
4. River Basin Analysis 363
4.1. Model Synthesis 363
4.2. Modelling Approach Using Optimization 364
4.3. Modelling Approach Using Simulation 365
4.4. Optimization and/or Simulation 368
4.5. Project Scheduling 368
5. Conclusions 371
6. References 371

12. Water Quality Modelling and Prediction 377
1. Introduction 377
2. Establishing Ambient Water Quality Standards 378
2.1. Water-Use Criteria 379
3. Water Quality Model Use 379
3.1. Model Selection Criteria 380
3.2. Model Chains 381
3.3. Model Data 382
4. Water Quality Model Processes 383
4.1. Mass-Balance Principles 384
4.1.1. Advective Transport 385
4.1.2. Dispersive Transport 385
4.1.3. Mass Transport by Advection and Dispersion 385
4.2. Steady-State Models 386
4.3. Design Streamflows for Water Quality 388
4.4. Temperature 389
4.5. Sources and Sinks 390
4.6. First-Order Constituents 390
4.7. Dissolved Oxygen 390
4.8. Nutrients and Eutrophication 393
4.9. Toxic Chemicals 396
4.9.1. Adsorbed and Dissolved Pollutants 396
4.9.2. Heavy Metals 398
4.9.3. Organic Micro-pollutants 399
4.9.4. Radioactive Substances 400
4.10. Sediments 400
4.10.1. Processes and Modelling Assumptions 401
4.10.2. Sedimentation 401
4.10.3. Resuspension 401
4.10.4. Burial 402
4.10.5. Bed Shear Stress 402
4.11. Lakes and Reservoirs 403
4.11.1. Downstream Characteristics 405
4.11.2. Lake Quality Models 406
4.11.3. Stratified Impoundments 407
5. An Algal Biomass Prediction Model 408
5.1. Nutrient Cycling 408
5.2. Mineralization of Detritus 408
5.3. Settling of Detritus and Inorganic Particulate Phosphorus 409
5.4. Resuspension of Detritus and Inorganic Particulate Phosphorus 409
5.5. The Nitrogen Cycle 409
5.5.1. Nitrification and Denitrification 409
5.5.2. Inorganic Nitrogen 410
5.6. Phosphorus Cycle 410
5.7. Silica Cycle 411
5.8. Summary of Nutrient Cycles 411
5.9. Algae Modelling 412
5.9.1. Algae Species Concentrations 412
5.9.2. Nutrient Recycling 413
5.9.3. Energy Limitation 413
5.9.4. Growth Limits 414
5.9.5. Mortality Limits 414
5.9.6. Oxygen-Related Processes 415
6. Simulation Methods 416
6.1. Numerical Accuracy 416
6.2. Traditional Approach 417
6.3. Backtracking Approach 418
6.4. Model Uncertainty 420
7. Conclusions: Implementing a Water Quality Management Policy 421
8. References 422

13. Urban Water Systems 427
1. Introduction 427
2. Drinking Water 428
2.1. Water Demand 428
2.2. Water Treatment 428
2.3. Water Distribution 430
2.3.1. Open Channel Networks 432

2.3.2. Pressure Pipe Networks 432
2.3.3. Water Quality 434
3. Wastewater 434
3.1. Wastewater Production 434
3.2. Sewer Networks 434
3.3. Wastewater Treatment 435
4. Urban Drainage 437
4.1. Rainfall 437
4.1.1. Time Series Versus Design Storms 437
4.1.2. Spatial-Temporal Distributions 438
4.1.3. Synthetic Rainfall 438
4.1.4. Design Rainfall 438
4.2. Runoff 439
4.2.1. Runoff Modelling 439
4.2.2. The Horton Infiltration Model 441
4.2.3. The US Soil Conservation Method (SCS) Model 442
4.2.4. Other Rainfall–Runoff Models 444
4.3. Surface Pollutant Loading and Washoff 445
4.3.1. Surface Loading 446
4.3.2. Surface Washoff 446
4.3.3. Stormwater Sewer and Pipe Flow 447
4.3.4. Sediment Transport 448
4.3.5. Structures and Special Flow Characteristics 448
4.4. Water Quality Impacts 448
4.4.1. Slime 448
4.4.2. Sediment 448
4.4.3. Pollution Impact on the Environment 448
4.4.4. Bacteriological and Pathogenic Factors 451
4.4.5. Oil and Toxic Contaminants 451
4.4.6. Suspended Solids 452
5. Urban Water System Modelling 452
5.1. Model Selection 452
5.2. Optimization 453
5.3. Simulation 455
6. Conclusions 456
7. References 457

## 14. A Synopsis 461

1. Meeting the Challenge 461
2. The Systems Approach to Planning and Management 461
2.1. Institutional Decision-Making 462
2.2. The Water Resources Systems 464
2.3. Planning and Management Modelling: A Review 465
3. Evaluating Modelling Success 466
4. Some Case Studies 467
4.1. Development of a Water Resources Management Strategy for Trinidad and Tobago 468
4.2. Transboundary Water Quality Management in the Danube Basin 470
4.3. South Yunnan Lakes Integrated Environmental Master Planning Project 473
4.4. River Basin Management and Institutional Support for Poland 475
4.5. Stormwater Management in The Hague in the Netherlands 476
5. Summary 478
6. References 478

## Appendix A: Natural System Processes and Interactions 480

1. Introduction 483
2. Rivers 483
2.1. River Corridor 484
2.1.1. Stream Channel Structure Equilibrium 485
2.1.2. Lateral Structure of Stream or River Corridors 486
2.1.3. Longitudinal Structure of Stream or River Corridors 487
2.2. Drainage Patterns 488
2.2.1. Sinuosity 489
2.2.2. Pools and Riffles 489
2.3. Vegetation in the Stream and River Corridors 489
2.4. The River Continuum Concept 490
2.5. Ecological Impacts of Flow 490
2.6. Geomorphology 490
2.6.1. Channel Classification 491
2.6.2. Channel Sediment Transport and Deposition 491
2.6.3. Channel Geometry 493
2.6.4. Channel Cross sections and Flow Velocities 494
2.6.5. Channel Bed Forms 495
2.6.6. Channel Planforms 495
2.6.7. Anthropogenic Factors 496
2.7. Water Quality 497
2.8. Aquatic Vegetation and Fauna 498
2.9. Ecological Connectivity and Width 500

2.10. Dynamic Equilibrium 501
2.11. Restoring Degraded Aquatic Systems 501
3. Lakes and Reservoirs 504
3.1. Natural Lakes 504
3.2. Constructed Reservoirs 505
3.3. Physical Characteristics 505
3.3.1. Shape and Morphometry 505
3.3.2. Water Quality 506
3.3.3. Downstream Characteristics 507
3.4. Management of Lakes and Reservoirs 508
3.5. Future Reservoir Development 510
4. Wetlands 510
4.1. Characteristics of Wetlands 511
4.1.1. Landscape Position 512
4.1.2. Soil Saturation and Fibre Content 512
4.1.3. Vegetation Density and Type 512
4.1.4. Interaction with Groundwater 513
4.1.5. Oxidation–Reduction 513
4.1.6. Hydrological Flux and Life Support 513
4.2. Biogeochemical Cycling and Storage 513
4.2.1. Nitrogen (N) 514
4.2.2. Phosphorus (P) 514
4.2.3. Carbon (C) 514
4.2.4. Sulphur (S) 514
4.2.5. Suspended Solids 514
4.2.6. Metals 515
4.3. Wetland Ecology 515
4.4. Wetland Functions 515
4.4.1. Water Quality and Hydrology 515
4.4.2. Flood Protection 516
4.4.3. Shoreline Erosion 516
4.4.4. Fish and Wildlife Habitat 516
4.4.5. Natural Products 516
4.4.6. Recreation and Aesthetics 516
5. Estuaries 516
5.1. Types of Estuaries 517
5.2. Boundaries of an Estuary 518
5.3. Upstream Catchment Areas 519
5.4. Water Movement 519
5.4.1. Ebb and Flood Tides 519
5.4.2. Tidal Excursion 520
5.4.3. Tidal Prism 520
5.4.4. Tidal Pumping 520
5.4.5. Gravitational Circulation 520
5.4.6. Wind-Driven Currents 521
5.5. Mixing Processes 521
5.5.1. Advection and Dispersion 522
5.5.2. Mixing 522
5.6. Salinity Movement 523
5.6.1. Mixing of Salt- and Freshwaters 523
5.6.2. Salinity Regimes 523
5.6.3. Variations due to Freshwater Flow 523
5.7. Sediment Movement 524
5.7.1. Sources of Sediment 524
5.7.2. Factors Affecting Sediment Movement 524
5.7.3. Wind Effects 525
5.7.4. Ocean Waves and Entrance Effects 525
5.7.5. Movement of Muds 526
5.7.6. Estuarine Turbidity Maximum 527
5.7.7. Biological Effects 527
5.8. Surface Pollutant Movement 528
5.9. Estuarine Food Webs and Habitats 528
5.9.1. Habitat Zones 529
5.10. Estuarine Services 531
5.11. Estuary Protection 531
5.12. Estuarine Restoration 533
5.13. Estuarine Management 533
5.13.1. Engineering Infrastructure 534
5.13.2. Nutrient Overloading 534
5.13.3. Pathogens 534
5.13.4. Toxic Chemicals 534
5.13.5. Habitat Loss and Degradation 534
5.13.6. Introduced Species 535
5.13.7. Alteration of Natural Flow Regimes 535
5.13.8. Declines in Fish and Wildlife Populations 535
6. Coasts 535
6.1. Coastal Zone Features and Processes 535
6.1.1. Water Waves 536
6.1.2. Tides and Water Levels 538
6.1.3. Coastal Sediment Transport 538
6.1.4. Barrier Islands 538
6.1.5. Tidal Deltas and Inlets 538
6.1.6. Beaches 538
6.1.7. Dunes 539
6.1.8. Longshore Currents 540
6.2. Coasts Under Stress 540
6.3. Management Issues 540
6.3.1. Beaches or Buildings 542
6.3.2. Groundwater 542

6.3.3. Sea Level Rise 542
6.3.4. Subsidence 543
6.3.5. Wastewater 544
6.3.6. Other Pollutants 544
6.3.7. Mining of Beach Materials 544
6.4. Management Measures 545
6.4.1. 'Conforming Use' 546
6.4.2. Structures 546
6.4.3. Artificial Beach Nourishment 547
7. Conclusion 548
8. References 549

## Appendix B: Monitoring and Adaptive Management 559

1. Introduction 559
2. System Status 561
2.1. System Status Indicators 562
3. Information Needs 562
3.1. Information Objectives and Priorities 563
4. Monitoring Plans 563
5. Adaptive Monitoring 564
5.1. Risk Assessments For Monitoring 564
5.2. Use of Models 565
6. Network Design 565
6.1. Site Selection 566
6.2. Sampling/Measurement Frequencies 566
6.3. Quality Control 566
6.4. Water Quantity Monitoring 567
6.5. Water Quality Monitoring 568
6.6. Ecological Monitoring 569
6.7. Early-Warning Stations 569
6.8. Effluent Monitoring 570
7. Data Sampling, Collection and Storage 570
7.1. Overview 570
7.2. Remote Sensing 571
7.2.1. Optical Remote Sensing for Water Quality 571
7.2.2. Applications in the North Sea 572
8. Data Analyses 572
9. Reporting Results 573
9.1. Trend Plots 573
9.2. Comparison Plots 573
9.3. Map Plots 576
10. Information Use: Adaptive Management 576
11. Summary 578
12. References 578

## Appendix C: Drought Management 581

1. Introduction 581
2. Drought Impacts 581
3. Defining Droughts 584
4. Causes of Droughts 585
4.1. Global Patterns 586
4.2. Teleconnections 588
4.3. Climate Change 588
4.4. Land Use 590
5. Drought Indices 590
5.1. Percent of Normal Indices 590
5.2. Standardized Precipitation Index 590
5.3. Palmer Drought Severity Index 591
5.4. Crop Moisture Index 592
5.5. Surface Water Supply Index 592
5.6. Reclamation Drought Index 593
5.7. Deciles 594
5.8. Method of Truncation 594
5.9. Water Availability Index 594
5.10. Days of Supply Remaining 595
6. Drought Triggers 596
7. Virtual Drought Exercises 596
8. Conclusion 598
9. References 599

## Appendix D: Flood Management 603

1. Introduction 603
2. Managing Floods in the Netherlands 605
2.1. Flood Frequency and Protection 605
2.2. The Rhine River Basin 605
2.3. Problems and Solutions 609
2.4. Managing Risk 609
2.4.1. Storage 610
2.4.2. Discharge-Increasing Measures 612
2.4.3. Green Rivers 614
2.4.4. Use of Existing Water Courses 615
2.4.5. The Overall Picture 615
2.5. Dealing With Uncertainties 615
2.6. Summary 617
3. Flood Management on the Mississippi 617
3.1. General History 619
3.2. Other Considerations 623
3.3. Interactions Among User Groups 624
3.4. Creating a Flood Management Strategy 626
3.5. The Role of the Government and NGOs 626

4. Flood Risk Reduction 627
4.1. Reservoir Flood Storage Capacity 627
4.2. Channel Capacity 630
4.3. Estimating Risk of Levee Failures 631
4.4. Annual Expected Damage From Levee Failure 633
4.4.1. Risk-Based Analyses 634
5. Decision Support and Prediction 635
5.1. Floodplain Modelling 636
5.2. Integrated 1D–2D Modelling 637
6. Conclusions 638
7. References 640

Appendix E: Project Planning and Analysis: Putting it All Together 644
1. Basic Concepts and Definitions 645
1.1. The Water Resources System 645
1.2. Functions of the Water Resources System 646
1.2.1. Subsistence Functions 646
1.2.2. Commercial Functions 646
1.2.3. Environmental Functions 647
1.2.4. Ecological Values 647
1.3. Policies, Strategies, Measures and Scenarios 647
1.4. Systems Analysis 648
2. Analytical Description of WRS 649
2.1. System Characteristics of the Natural Resources System 650
2.1.1. System Boundaries 650
2.1.2. Physical, Chemical and Biological Characteristics 650
2.1.3. Control Variables: Possible Measures 651
2.2. System Characteristics of the Socio-Economic System 651
2.2.1. System Boundaries 651
2.2.2. System Elements and System Parameters 651
2.2.3. Control Variables: Possible Measures 652
2.3. System Characteristics of the Administrative and Institutional System 652
2.3.1. System Elements 652
2.3.2. Control Variables: Possible Measures 652
3. Analytical Framework: Phases of Analysis 652
4. Inception Phase 654
4.1. Initial Analysis 655
4.1.1. Inventory of Characteristics, Developments and Policies 655
4.1.2. Problem Analysis 655
4.1.3. Objectives and Criteria 656
4.1.4. Data Availability 656
4.2. Specification of the Approach 657
4.2.1. Analysis Steps 657
4.2.2. Delineation of System 657
4.2.3. Computational Framework 658
4.2.4. Analysis Conditions 659
4.2.5. Work Plan 660
4.3. Inception Report 660
4.4. Communication with Decision-Makers and Stakeholders 661
5. Development Phase 661
5.1. Model Development and Data Collection 661
5.1.1. Analysis of the Natural Resources System (NRS) 661
5.1.2. Analysis of the Socio-Economic System (SES) 664
5.1.3. Analysis of the Administrative and Institutional System (AIS) 666
5.1.4. Integration into a Computational Framework 667
5.2. Preliminary Analysis 668
5.2.1. Base Case Analysis 669
5.2.2. Bottleneck (Reference Case) Analysis 669
5.2.3. Identification and Screening of Measures 669
5.2.4. Finalization of the Computational Framework 669
6. Selection Phase 670
6.1. Strategy Design and Impact Assessment 670
6.2. Evaluation of Alternative Strategies 671
6.3. Scenario and Sensitivity Analysis 672
6.4. Presentation of Results 672
7. Conclusions 672

Index 677

# Introduction

Water resources are special. In their natural states they are beautiful. People like to live and vacation near rivers, lakes and coasts. Water is also powerful. Water can erode rock, alter existing landscapes and form new ones. Life on this planet depends on water. Most of our economic activities consume water. All of the food we grow, process and eat requires water. Much of our waste is transported and assimilated by water. The importance of water to our well-being is beyond question. Our dependence on water will last forever.

So, what is the problem? The answer is simply that water, although plentiful, is not distributed as we might wish. There is often too much or too little, or what exists is too polluted or too expensive. A further problem is that the overall water situation is likely to further deteriorate as a result of global changes. This is a result not only of climatic change but also of other global change drivers such as population growth, land use changes, urbanization and migration from rural to urban areas, all of which will pose challenges never before seen. Water obviously connects all these areas and any change in these drivers has an impact on it. Water has its own dynamics that are fairly non-linear. For example, while population growth in the twentieth century increased three-fold – from 1.8 billion to 6 billion people – water withdrawal during the same period increased six-fold! That is clearly unsustainable. Freshwater, although a renewable resource, is finite and is very vulnerable. If one considers all the water on Earth, 97.5% is located in the seas and oceans and what is available in rivers, lakes and reservoirs for immediate human consumption comprises no more than a mere 0.007 per cent of the total. This is indeed very limited and on average is roughly equivalent to 42,000 cubic kilometres per year.

If one looks at the past thirty years only in terms of reduction in per capita water availability in a year the picture is even more disturbing. While in 1975 availability stood at around 13,000 cubic metres per person per year, it has now dropped to 6,000 cubic metres; meanwhile water quality has also severely deteriorated. While this cannot be extrapolated in any meaningful manner, it nevertheless indicates the seriousness of the situation. This will likely be further exacerbated by the expected impacts of climate change. Although as yet unproven to the required rigorous standards of scientific accuracy, increasing empirical evidence indicates that the hydrological cycle is accelerating while the amount of water at a given moment in time is remains the same. If this acceleration hypothesis is true then it will cause an increase in the frequency and magnitude of flooding. At the other end of the spectrum, the prevailing laws of continuity mean that the severity and duration of drought will also increase. These increased risks are likely to have serious regional implications. Early simulation studies, carried out by IHP, suggest that wet areas will become even more humid while dry areas will become increasingly arid. This will not occur overnight; similarly, appropriate countermeasures will need time to establish policies that integrate the technical and social issues in a way that takes appropriate consideration of the cultural context.

Tremendous efforts and political will are needed to achieve the two water related Millennium Development Goals (MDGs), that is, to halve the number of human beings who have no access to safe drinking water and adequate sanitation facilities respectively, by 2015. In the case of drinking water, we have 1.2 billion fellow human beings that have no access to safe drinking water, while in the case of sanitation, the figure is 2.4 billion.

The substantial growth of human populations – especially as half of humanity already lives in urban areas – and the consequent expansion of agricultural and industrial activities with a high water demand, have only served to increase problems of water availability, quality – and in many regions – waterborne disease. There is now an increasing urgency in the UN system to protect water resources through better management. Data on the scale of deforestation with subsequent land use conversion, soil erosion, desertification, urban sprawl, loss of genetic diversity, climate change and the precariousness of food production through irrigation, all reveal the growing seriousness of the problem. We have been forced to recognize that society's activities can no longer continue unchecked without causing serious damage to the very environment and ecosystems we depend upon for our survival. This is especially critical in water scarce regions, many of which are found in the developing world and are dependent on water from aquifers that are not being recharged as fast as their water is being withdrawn and consumed. Such practices are clearly not sustainable.

Proper water resources management requires consideration of both supply and demand. The mismatch of supply and demand over time and space has motivated the development of much of the water resources infrastructure that is in place today. Some parts of the globe witness regular flooding as a result of monsoons and torrential downpours, while other areas suffer from the worsening of already chronic water shortages. These conditions are often aggravated by the increasing discharge of pollutants resulting in a severe decline in water quality.

The goal of sustainable water management is to promote water use in such a way that society's needs are both met to the extent possible now and in the future. This involves protecting and conserving water resources that will be needed for future generations. UNESCO's International Hydrological Programme (IHP) addresses these short- and long-term goals by advancing our understanding of the physical and social processes affecting the globe's water resources and integrating this knowledge into water resources management. This book describes the kinds of problems water managers can and do face and the types of models and methods one can use to define and evaluate alternative development plans and management policies. The information derived from these models and methods can help inform stakeholders and decision-makers alike in their search for sustainable solutions to water management problems. The successful application of these tools requires collaboration among natural and social scientists and those in the affected regions, taking into account not only the water-related problems but also the social, cultural and environmental values.

On behalf of UNESCO it gives me great pleasure to introduce this book. It provides a thorough introduction to the many aspects and dimensions of water resources management and presents practical approaches for analysing problems and identifying ways of developing and managing water resources systems in a changing and uncertain world. Given the practical and academic experience of the authors and the contributions they have made to our profession, I am confident that this book will become a valuable asset to those involved in water resources planning and management. I wish to extend our deepest thanks to Professors Pete Loucks and Eelco van Beek for offering their time, efforts and outstanding experience, which is summarized in this book for the benefit of the growing community of water professionals.

**András Szöllösi-Nagy**
Deputy Assistant Director-General, UNESCO
Secretary, International Hydrological Programme

# 1. Water Resources Planning and Management: An Overview


1. Introduction 3

2. Planning and Management Issues: Some Case Studies 4
   - 2.1. Kurds Seek Land, Turks Want Water 4
   - 2.2. Sharing the Water of the Jordan River Basin: Is There a Way? 6
   - 2.3 Mending the 'Mighty and Muddy' Missouri 7
   - 2.4. The Endangered Salmon 7
   - 2.5. The Yellow River: How to Keep the Water Flowing 9
   - 2.6. Lake Source Cooling: Aid to Environment or Threat to Lake? 10
   - 2.7. Managing Water in the Florida Everglades 11
   - 2.8. Restoration of Europe's Rivers and Seas 13
     - 2.8.1. The Rhine 13
     - 2.8.2. The Danube 14
     - 2.8.3. The North and Baltic Seas 15
   - 2.9. Egypt and the Nile: Limits to Agricultural Growth 16
   - 2.10. Damming the Mekong 16

3. So, Why Plan, Why Manage? 18
   - 3.1. Too Little Water 20
   - 3.2. Too Much Water 20
   - 3.3. Polluted Water 21
   - 3.4. Degradation of Aquatic and Riparian Ecosystems 21
   - 3.5. Other Planning and Management Issues 21

4. System Components, Planning Scales and Sustainability 22
   - 4.1. Spatial Scales for Planning and Management 22
   - 4.2. Temporal Scales for Planning and Management 23
   - 4.3. Sustainability 23

5. Planning and Management 24
   - 5.1. Approaches 24
     - 5.1.1. Top-Down Planning and Management 25
     - 5.1.2. Bottom-Up Planning and Management 25
     - 5.1.3. Integrated Water Resources Management 26
   - 5.2. Planning and Management Aspects 26
     - 5.2.1. Technical Aspects 26
     - 5.2.2. Economic and Financial Aspects 27
     - 5.2.3. Institutional Aspects 28
   - 5.3. Analysis for Planning and Management 28
   - 5.4. Models for Impact Prediction and Evaluation 30
   - 5.5. Shared-Vision Modelling 31
   - 5.6. Adaptive Integrated Policies 31
   - 5.7. Post-Planning and Management Issues 32

6. Meeting the Planning and Management Challenges: A Summary 32

7. References 34

# 1 Water Resources Planning and Management: An Overview

Water resource systems have benefited both people and their economies for many centuries. The services provided by such systems are multiple. Yet in many regions, water resource systems are not able to meet the demands, or even the basic needs, for clean fresh water, nor can they support and maintain resilient biodiverse ecosystems. Typical causes of such failures include degraded infrastructures, excessive withdrawals of river flows, pollution from industrial and agricultural activities, eutrophication from excessive nutrient loads, salinization from irrigation return flows, infestations of exotic plants and animals, excessive fish harvesting, floodplain and habitat alteration from development activities, and changes in water and sediment flow regimes. Inadequate water resource systems reflect failures in planning, management and decision-making – and at levels broader than water. Planning, developing and managing water resource systems to ensure adequate, inexpensive and sustainable supplies and qualities of water for both humans and natural ecosystems can only be successful if such activities address the causal socio-economic factors, such as inadequate education, population pressures and poverty.

## 1. Introduction

Over the centuries, surface and ground waters have been a source of water supplies for agricultural, municipal and industrial consumers. Rivers have provided hydroelectric energy and inexpensive ways of transporting bulk cargo between different ports along their banks, as well as water-based recreational opportunities, and have been a source of water for wildlife and its habitat. They have also served as a means of transporting and transforming waste products that are discharged into them. The quantity and quality regimes of streams and rivers have been a major factor in governing the type, health and biodiversity of riparian and aquatic ecosystems. Floodplains have provided fertile lands for agricultural production and relatively flat lands for roads, railways and commercial and industrial complexes. In addition to the economic benefits that can be derived from rivers and their floodplains, the aesthetic beauty of most natural rivers has made lands adjacent to them attractive sites for residential and recreational development. Rivers and their floodplains have generated and, if managed properly, can continue to generate substantial economic, environmental and social benefits for their inhabitants.

Human activities undertaken to increase the benefits obtained from rivers and their floodplains may also increase the potential for costs and damage when the river is experiencing rare or extreme flow conditions, such as during periods of drought, floods and heavy pollution. These costs and impacts are economic, environmental and social in nature and result from a mismatch between what humans expect or demand, and what nature (and occasionally our own activities) offers or supplies. Human activities tend to be based on the 'usual or normal' range of river flow conditions. Rare or 'extreme' flow or water quality conditions outside these normal ranges will continue to occur, and possibly with increasing frequency as climate change experts suggest. River-dependent, human activities that cannot adjust to these occasional extreme conditions will incur losses.

The planning of human activities involving rivers and their floodplains must consider certain hydrological facts. One of these facts is that flows and storage volumes vary over space and time. They are also finite. There are limits to the amounts of water that can be withdrawn from surface and groundwater bodies. There are also limits to the amounts of potential pollutants that can be discharged into them without causing damage. Once these limits are exceeded, the concentrations of pollutants in these waters may reduce or even eliminate the benefits that could be obtained from other uses of the resource.

Water resources professionals have learned how to plan, design, build and operate structures that, together with non-structural measures, increase the benefits people can obtain from the water resources contained in rivers and their drainage basins. However, there is a limit to the services one can expect from these resources. Rivers, estuaries and coastal zones under stress from overdevelopment and overuse cannot reliably meet the expectations of those depending on them. How can these renewable yet finite resources best be managed and used? How can this be accomplished in an environment of uncertain supplies and uncertain and increasing demands, and consequently of increasing conflicts among individuals having different interests in the management of a river and its basin? The central purpose of water resources planning and management activities is to address and, if possible, answer these questions. These issues have scientific, technical, political (institutional) and social dimensions and thus, so must water resources planning processes and their products.

River basin, estuarine and coastal zone managers – those responsible for managing the resources in those areas – are expected to manage them effectively and efficiently, meeting the demands or expectations of all users and reconciling divergent needs. This is no small task, especially as demands increase, as the variability of hydrological and hydraulic processes becomes more pronounced, and as stakeholder measures of system performance increase in number and complexity. The focus or goal is no longer simply to maximize net economic benefits while ensuring the equitable distribution of those benefits. There are also environmental and ecological goals to consider. Rarely are management questions one-dimensional, such as: 'How can we provide more high-quality water to irrigation areas in the basin at acceptable costs?' Now added to that question is how those withdrawals would affect the downstream water quantity and quality regimes, and in turn the riparian and aquatic ecosystems. To address such 'what if' questions requires the integration of a variety of sciences and technologies with people and their institutions.

Problems and opportunities change over time. Just as the goals of managing and using water change over time, so do the processes of planning to meet these changing goals. Planning processes evolve not only to meet new demands, expectations and objectives, but also in response to new perceptions of how to plan more effectively.

This book is about how quantitative analysis, and in particular computer models, can support and improve water resources planning and management. This first chapter attempts to review some of the issues involved. It provides the context and motivation for the chapters that follow, which describe in more detail our understanding of 'how to plan' and 'how to manage' and how computer-based programs and models can assist those involved in these activities. Additional information is available in many of the references listed at the end of each chapter.

## 2. Planning and Management Issues: Some Case Studies

Managing water resources certainly requires knowledge of the relevant physical sciences and technology. But at least as important, if not more so, are the multiple institutional, social or political issues confronting water resources planners and managers. The following brief descriptions of some water resources planning and management studies at various geographic scales illustrate some of these issues.

### 2.1. Kurds Seek Land, Turks Want Water

The Tigris and Euphrates Rivers (Figure 1.1) in the Middle East created the 'Fertile Crescent' where some of the first civilizations emerged. Today their waters are


**Figure 1.1.** The Tigris and Euphrates Rivers in Turkey, northern Syria and Iraq.

**Figure 1.2.** Ataturk Dam on the Euphrates River in Turkey (DSI).

critical resources, politically as well as geographically. In one of the world's largest public works undertakings, Turkey is spending over $30 billion in what is called the Great Anatolia Project (GAP), a complex of 22 reservoirs and 19 hydroelectric plants. Its centrepiece, the Ataturk Dam (Figure 1.2) on the Euphrates River, is already completed. In the lake formed behind the dam, sailing and swimming competitions are being held on a spot where, for centuries, there was little more than desert (Figure 1.3).

**Figure 1.3.** Water sports on the Ataturk Reservoir on the Euphrates River in Turkey (DSI).

When the project is completed it is expected to increase the amount of irrigated land in Turkey by 40% and provide up to a quarter of the country's electric power needs. Planners hope this can improve the standard of living of six million of Turkey's poorest people, most of them Kurds, and thus undercut the appeal of revolutionary separatism. It will also reduce the amount of water Syria and Iraq believe they need – water that Turkey fears might ultimately be used for anti-Turkish causes.

The region of Turkey where Kurd's predominate is more or less the same region covered by the Great Anatolia Project, encompassing an area about the size of Austria. Giving that region autonomy by placing it under Kurdish self-rule could weaken the Central Government's control over the water resources that it recognizes as a keystone of its future power.

In other ways also, Turkish leaders are using their water as a tool of foreign as well as domestic policy. Among their most ambitious projects considered is a fifty-mile undersea pipeline to carry water from Turkey to the parched Turkish enclave on northern Cyprus. The pipeline, if actually built, will carry more water than northern Cyprus can use. Foreign mediators, frustrated by their inability to break the political deadlock on Cyprus, are hoping that the excess water can be sold to the ethnic Greek republic on the southern part of the island as a way of promoting peace.

## 2.2. Sharing the Water of the Jordan River Basin: Is There a Way?

A growing population – approximately 12 million people – and intense economic development in the Jordan River Basin (Figure 1.4) are placing heavy demands on its scarce freshwater resources. Though the largely arid region receives less than 250 millimetres of rainfall each year, total water use for agricultural and economic activities has been steadily increasing. This and encroaching urban development have degraded many sources of high-quality water in the region.

The combined diversions by the riparian water users have changed the river in its lower course into little better than a sewage ditch. Of the 1.3 billion cubic metres (mcm or $10^6$ $m^3$) of water that flowed into the Dead Sea in the 1950s, only a small fraction remains at present. In normal years the flow downstream from Lake Tiberias (also called the Sea of Galilee or Lake Kinneret) is some 60 mcm – about 10% of the natural discharge in this section. It mostly consists of saline springs and sewage water. These flows are then joined by what remains of the Yarmouk, by some irrigation return flows and by winter runoff, adding up to an annual total of 200 to 300 mcm. This water is unsuitable for irrigation in both quantity and quality, nor does it sufficiently supply natural systems. The salinity of the Jordan River reaches up to 2,000 parts per million (ppm) in the lowest section, which renders it unfit for crop irrigation. Only in flood years is fresh water released into the lower Jordan Valley.


**Figure 1.4.** The Jordan River between Israel and Jordan.

One result of this increased pressure on freshwater resources is the deterioration of the region's wetlands, which are important for water purification and flood and erosion control. As agricultural activity expands, wetlands are being drained, and rivers, aquifers, lakes and streams are being polluted with runoff containing fertilizers and pesticides. Reversing these trends by preserving natural ecosystems is essential to the future availability of fresh water in the region.

To ensure that an adequate supply of fresh, high-quality water is available for future generations, Israel, Jordan and the Palestinian Authority will have to work together to preserve aquatic ecosystems (National Research Council, 1999). Without these natural ecosystems, it will be difficult and expensive to sustain high-quality water supplies. The role of ecosystems in sustaining water resources has largely been overlooked in the context of the region's water provision. Vegetation controls storm runoff, filters polluted water and reduces erosion and the amount of sediment that makes its way into water supplies. Streams assimilate wastewater, lakes store clean water, and surface waters provide habitats for many plants and animals.

The Jordan River Basin, like most river basins, should be evaluated and managed as a whole to permit the comprehensive assessment of the effects of water management options on wetlands, lakes, the lower river and the Dead Sea coasts. Damage to ecosystems and loss of animal and plant species should be weighed against the potential benefits of developing land and creating new water resources. For example, large river-management projects that divert water to dry areas have promoted intensive year-round farming and urban development, but available river water is declining and becoming increasingly polluted. Attempting to meet current demands solely by withdrawing more ground and surface water could result in widespread environmental degradation and depletion of freshwater resources.

There are policies that, if implemented, could help preserve the capacity of the Jordan River to meet future demands. Most of the options relate to improving the efficiency of water use: that is, they involve conservation and better use of proven technologies. Also being considered are policies that emphasize economic efficiency and reduce overall water use. Charging higher rates for water use in peak periods and surcharges for excessive use, would encourage conservation. In addition, new sources of fresh water can be obtained by capturing rainfall through rooftop cisterns, catchment systems and storage ponds.

Thus there are alternatives to a steady deterioration of the water resources of the Jordan Basin. They will require coordination and cooperation among all those living in the basin. Will this be possible?

## 2.3. Mending the 'Mighty and Muddy' Missouri

Nearly two centuries after an epic expedition through the western United States in search of a northwest river passage to the Pacific Ocean, there is little enchantment left to the Missouri River. Shown in Figure 1.5, it has been dammed, dyked and dredged since the 1930s to control floods and float cargo barges. The river nicknamed the 'Mighty Missouri' and the 'Big Muddy' by its explorers is today neither mighty nor very muddy. The conservation group American Rivers perennially lists the Missouri among the United States' ten most endangered rivers.

Its wilder upper reaches are losing their cottonwood trees to dam operations and cattle that trample seedlings along the river's banks. In its vast middle are multiple dams that hold back floods, generate power and provide pools for boats and anglers.

Its lower third is a narrow canal sometimes called 'the Ditch' that is deep enough for commercial tow-boats. Some of the river's banks are armoured with rock and concrete retaining walls that protect half a million acres of farm fields from flooding. Once those floods produced and maintained marshlands and side streams – habitats for a wide range of wildlife. Without these habitats, many wild species are unable to thrive, or in some cases even survive.

Changes to restore at least some of the Missouri to a more natural state are being implemented. Protection of fish and wildlife habitat has been added to the list of objectives to be achieved by the government agencies managing the Missouri. The needs of wildlife are now seen to be as important as other competing interests on the river, including navigation and flood control. This is in reaction, in part, to the booming $115 million-a-year outdoor recreation industry. Just how much more emphasis will be given to these back-to-nature goals depends on whether the Missouri River Basin Association, an organization representing eight states and twenty-eight Native American tribes, can reach a compromise with the traditional downstream uses of the river.


**Figure 1.5. Major rivers in the continental United States.**

## 2.4. The Endangered Salmon

Greater Seattle in the northwestern US state of Washington may be best known around the world for its software and aviation industry, but residents know it for something less flashy: its dwindling stock of wild salmon (see Figure 1.6). The Federal Government has placed seven types of salmon and two types of trout on its list of threatened or endangered species. Saving the fish from extinction will require sacrifices and could slow development in one of the fastest-growing regions of the United States.

Before the Columbia River and its tributaries in the northwestern United States were blocked with dozens of dams, about 10 to 16 million salmon made the annual run back up to their spawning grounds. In 1996, a little less than a million did. But the economy of the Northwest depends on the dams and locks that have been built in the Columbia to provide cheap hydropower production and navigation.

**Figure 1.6.** A salmon swimming upstream (US Fish and Wildlife Service, Pacific Region).

For a long time, engineers tried to jury-rig the system so that fish passage would be possible. It has not worked all that well. Still too many young fish enter the hydropower turbines on their way down the river. Now, as the debate over whether or not to remove some dams takes place, fish are caught and carried by truck around the turbines. The costs of keeping these salmon alive, if not completely happy, are enormous.

Over a dozen national and regional environmental organizations have joined together to bring back salmon and steelhead by modifying or partially dismantling five federal dams on the Columbia and Snake Rivers. Partial removal of the four dams on the lower Snake River in Washington State and lowering the reservoir behind John Day Dam on the Columbia bordering Oregon and Washington (see Figure 1.7) should help restore over 300 km of vital river habitat. Running the rivers in a more


**Figure 1.7.** The Snake and Columbia River reservoirs identified by the Columbia and Snake Rivers Campaign for modification or dismantling to permit salmon passage.

natural way may return salmon and steelhead to the harvestable levels of the 1960s before the dams were built.

Dismantling part of the four Lower Snake dams will leave most of each dam whole. Only the dirt bank connecting the dam to the riverbank will be removed. The concrete portion of the dam will remain in place, allowing the river to flow around it. The process is reversible and, the Columbia and Snake Rivers Campaign argues, it will actually save taxpayers money in planned dam maintenance by eliminating subsidies to shipping industries and agribusinesses, and by ending current salmon recovery measures that are costly. Only partially removing the four Lower Snake River dams and modifying John Day Dam will restore rivers, save salmon and return balance to the Northwest's major rivers.

## 2.5. The Yellow River: How to Keep the Water Flowing

The Yellow River is one of the most challenging in the world from the point of view of water and sediment management. Under conditions of normal and low flow, the water is used for irrigation, drinking and industry to such an extent that the lower reach runs dry during many days each year. Under high-flow conditions, the river is heavily laden with very fine sediment originating from the Löss Plateau, to the extent that a hyperconcentrated flow occurs. Through the ages the high sediment load has resulted in the building-out of a large delta in the Bohai Sea and a systematic increase of the large-scale river slope. Both have led to what is now called the 'suspended river': the riverbed of the lower reach is at points some 10 metres above the adjacent land, with dramatic effects if dyke breaching were to occur.

The Yellow River basin is already a very water-scarce region. The rapid socio-economic development in China is putting the basin under even more pressure. Agricultural, industrial and population growth will further increase the demand for water. Pollution has reached threatening levels. The Chinese government, in particular the Yellow River Conservancy Commission (YRCC), has embarked on an ambitious program to control the river and regulate the flows. Their most recent accomplishment is the construction of the Xiaolangdi Dam, which will


**Figure 1.8.** The Yellow River Basin.

control water and sediment just before the river enters the flat lower reach. This controlling includes a concentrated release of high volumes of water to flush the sediment out to sea.

In the delta of the Yellow River, fresh water wetlands have developed with a dynamic and unique ecosystem of valuable plant species and (transmigratory) birds. The decreased and sometimes zero flow in the river is threatening this ecosystem. To protect it, the YRCC has started to release additional water from the Xiaolangdi dam to 'supply' these wetlands with water during dry periods. The water demand of the wetlands is in direct competition with the agricultural and industrial demands upstream, and there have been massive complaints about this 'waste' of valuable water. Solving this issue and agreeing upon an acceptable distribution over users and regions is a nearly impossible task, considering also that the river crosses nine rather autonomous provinces.

How can water be kept flowing in the Yellow River basin? Under high-flow conditions the sediment has to be flushed out of the basin to prevent further build-up of the suspended river. Under low-flow conditions water has to be supplied to the wetlands. In both cases the water is seen as lost for what many consider to be its main function: to support the socio-economic development of the region.

## 2.6. Lake Source Cooling: Aid to Environment or Threat to Lake?

It seems an environmentalist's dream: a cost-effective system that can cool some 10 million square feet of high school and university buildings simply by pumping cold water from the depths of a nearby lake (Figure 1.9), without the emission of chlorofluorocarbons (the refrigerants that can destroy protective ozone in the atmosphere) and at a cost substantially smaller than for conventional air conditioners. The water is returned to the lake, with a few added calories.

However, a group of local opponents insists that Cornell University's $55-million lake-source-cooling plan, which has replaced its aging air conditioners, is actually an environmental threat. They believe it could foster algal blooms. Pointing to five years of studies, thousands of pages of data, and more than a dozen permits from local and state agencies, Cornell's consultants say the system could actually improve conditions in the lake. Yet another benefit, they say, is that the system would reduce Cornell's contribution to global warming by reducing the need to burn coal to generate electricity.

**Figure 1.9.** The cold deep waters of Lake Cayuga are being used to cool the buildings of a local school and university (Ithaca City Environmental Laboratory).

For the most part, government officials agree. But a small determined coalition of critics from the local community argue over the expected environmental impacts, and over the process of getting the required local, state and federal permits approved. This is in spite of the fact that the planning process, which took over five years, requested and involved the participation of all interested stakeholders from the very beginning. Even the local chapter of the Sierra Club and biology professors at other universities have endorsed the project. However, in almost every project where the environmental impacts are uncertain, there will be debates among scientists as well as among stakeholders. In addition, a significant segment of society distrusts scientists anyway. 'This is a major societal problem,' wrote a professor and expert in the dynamics of lakes. 'A scientist says X and someone else says Y and you've got chaos. In reality, we are the problem. Every time we flush our toilets, fertilize our lawns, gardens and fields, or wash our cars, we contribute to the nutrient loading of the lake.'

The project has now been operating for over five years, and so far no adverse environmental effects have been noticed at any of the many monitoring sites.

## 2.7. Managing Water in the Florida Everglades

The Florida Everglades (Figure 1.10) is the largest single wetland in the continental United States. In the mid-1800s it covered a little over 3.6 million ha, but since that time the historical Everglades has been drained and half of the area is now devoted to agriculture and urban development. The remaining wetland areas have been altered by human disturbances both around and within them. Water has been diverted for human uses, flows have been lowered to protect against floods, nutrient supplies to the wetlands from runoff from agricultural fields and urban areas have increased, and invasions of non-native or otherwise uncommon plants and animals have out-competed native species. Populations of wading birds (including some endangered species) have declined by 85 to 90% in the last half-century, and many species of South Florida's mammals, birds, reptiles, amphibians and plants are either threatened or endangered.

The present management system of canals, pumps, and levees (Figure 1.11) will not be able to provide adequate water supplies or sufficient flood protection to agricultural and urban areas, let alone support the natural (but damaged) ecosystems in the remaining wetlands. The system is not sustainable. Problems in the greater Everglades ecosystem relate to both water quality and quantity, including the spatial and temporal distribution of water depths, flows and flooding durations (called hydroperiods). Issues arise due to variations in

**Figure 1.10.** Scenes of the Everglades in southern Florida (South Florida Water Management District).

**Figure 1.11.** Pump station on a drainage canal in southern Florida (South Florida Water Management District).


**Figure 1.12.** Europe's major rivers and seas.

the natural/historical hydrological regime, degraded water quality and the sprawl from fast-growing urban areas.

To meet the needs of the burgeoning population and increasing agricultural demands for water, and to begin the restoration of the Everglades' aquatic ecosystem to a more natural state, an ambitious plan has been developed by the US Army Corps of Engineers (USACE) and its local sponsor, the South Florida Water Management District. The proposed Corps plan is estimated to cost over $8 billion. The plan and its Environmental Impact Statement (EIS) have received input from many government agencies and non-governmental organizations, as well as from the public at large.

The plan to restore the Everglades is ambitious and comprehensive, involving the change of the current hydrological regime in the remnant of the Everglades to one that resembles a more natural one, the re-establishment of marshes and wetlands, the implementation of agricultural best-management practices, the enhancement of wildlife and recreation areas, and the distribution of provisions for water supply and flood control to the urban population, agriculture and industry.

Planning for and implementing the restoration effort requires application of state-of-the-art large systems analysis concepts, hydrological and hydroecological data and models incorporated within decision support systems, integration of social sciences, and monitoring for planning and evaluation of performance in an adaptive management context. These large, complex challenges of the greater Everglades restoration effort demand the most advanced, interdisciplinary and scientifically-sound analysis capabilities available. They also require the political will to make compromises and to put up with lawsuits by anyone who may be disadvantaged by some restoration measure.

Who pays for all this? Both the taxpayers of Florida, and the taxpayers of the United States.

## 2.8. Restoration of Europe's Rivers and Seas

### 2.8.1. The Rhine

The map of Figure 1.13 shows the areas of the nine countries that are part of river Rhine basin. In the Dutch area of the Rhine basin, water is partly routed northward through the Ijssel and westward through the highly interconnected river systems of the Rhine, Meuse and Waal. About 55 million people live in the Rhine River basin and about 20 million of those people drink the river water.

In the mid-1970s, some called the Rhine the most romantic sewer in Europe. In November 1986, a chemical spill degraded much of the upper Rhine's aquatic ecosystem. This damaging event was reported worldwide. The Rhine was again world news in the first two months of 1995 when its water level reached a height that occurs on average once in a century. In the Netherlands, some 200,000 people, 1,400,000 pigs and cows and 1,000,000 chickens had to be evacuated. During the last two months of the same year there was hardly enough water in the

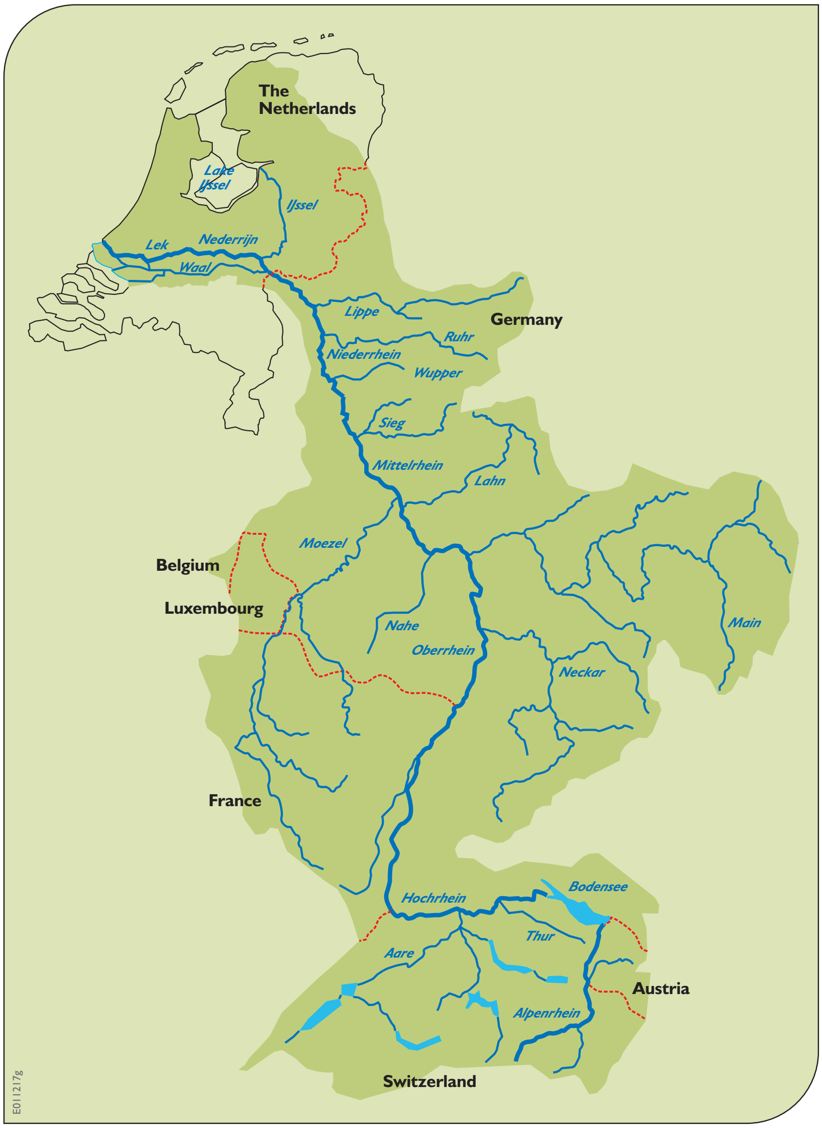



**Figure 1.13.** The Rhine River basin of western Europe and its extent in the Netherlands.

Rhine for navigation. It is fair to say these events have focused increased attention on what needs to be done to 'restore' and protect the Rhine.

To address just how to restore the Rhine, it is useful to look at what has been happening to the river during the past 150 years. The Rhine, the only river connecting the Alps with the North Sea, was originally a natural watercourse. To obtain greater economic benefits from the river, it was engineered for navigation, hydropower, water supply and flood protection. Floodplains, now 'protected' from floods, provided increased land areas suitable for development. The main stream of the Rhine is now considerably shorter, narrower and deeper than it was originally.

From an economic development point of view, the engineering works implemented in the river and its basin worked. The Rhine basin is now one of the most industrialized regions in the world and is characterized by intensive industrial and agricultural activities: it contains some 20% of the world's chemical industry. The river is reportedly the busiest shipping waterway in the world, containing long canals with regulated water levels, connecting the Rhine and its tributaries with the rivers of almost all the surrounding river basins, including the Danube. This provides water transport to and from the North and Black Seas.

From an environmental and ecological viewpoint, and from the viewpoint of flood control as well, the economic development that has taken place over the past two centuries has not worked perfectly. The concerns aroused by the recent toxic spill and floods, and from a generally increasing interest by the inhabitants of the basin in environmental and ecosystem restoration and the preservation of natural beauty, have resulted in basinwide efforts to rehabilitate the basin to a more 'living' sustainable entity.

A Rhine Action Programme has been created to revive the ecosystem. The goal of that program is the revival of the main stream as the backbone of the ecosystem, particularly for migratory fish, and the protection, maintenance and revival of ecologically important areas along the Rhine. Implemented in the 1990s, the plan was given the name 'Salmon 2000,' since the return of salmon to the Rhine is seen as a symbol of ecological revival. A healthy salmon population will need to swim throughout the river length. This will be a challenge, as no one pretends that the engineering works that provide navigation and hydropower benefits, but which also inhibit fish passage, are no longer needed or desired.

### 2.8.2. The Danube

The Danube River (shown in Figure 1.14) is in the heartland of central Europe. Its basin includes large parts of the territories of thirteen countries. It additionally receives


**Figure 1.14.** The Danube River in central Europe.

runoff from small catchments located in five other countries. About 85 million people live in the basin. This river encompasses greater political, economic and social variations than arguably any other river basin in Europe.

The river discharges into the Black Sea. The Danube delta and the banks of the Black Sea have been designated a biosphere reserve by UNESCO. Over half of the delta has been declared a 'wet zone of international significance.' Throughout its length the Danube provides a vital resource for drainage, communications, transport, power generation, fishing, recreation and tourism. It is considered to be an ecosystem of irreplaceable environmental value.

More than forty dams and large barrages, plus over 500 smaller reservoirs have been constructed on the main Danube River and its tributaries. Flood-control dykes confine most of the length of the main stem of the Danube River and the major tributaries. Over the last fifty years natural alluvial floodplain areas have declined from about 26,000 $km^2$ to about 6,000 $km^2$.

There are also significant reaches with river training works and river diversion structures. These structures trap nutrients and sediment in the reservoirs, which causes changes in downstream flow and sediment transport regimes that reduce the ecosystems' habitats both longitudinally and transversely and decrease the efficiency of natural purification processes. Thus, while these engineered facilities provide important opportunities for the control and use of the river's resources, they also illustrate the difficulties of balancing these important economic activities with environmentally sound and sustainable management.

The environmental quality of the Danube River is also under intense pressure from a diverse range of human activities, including point source and non-point source agricultural, industrial and municipal wastes. Because of the poor water quality (sometimes affecting human health), the riparian countries of the basin have been participating in environmental management activities on regional, national and local levels for several decades. All Danube countries signed a formal Convention on Cooperation for the Protection and Sustainable Use of the Danube River in June 1994. The countries have agreed to take 'all appropriate legal, administrative and technical measures to improve the current environmental and water quality conditions of the Danube River and of the waters in its catchment area, and to prevent and reduce as much as possible the adverse impacts and changes occurring or likely to be caused'.

### 2.8.3. The North and Baltic Seas

The North and Baltic Seas (shown in Figure 1.12) are the most densely navigated seas in the world. Besides shipping, military and recreational uses, there is an offshore oil and gas industry, and telephone cables cover the seabed. The seas are rich and productive, with resources that include not only fish but also crucial minerals (in addition to oil) such as gas, sand and gravel. These resources and activities play major roles in the economies of the surrounding countries.

Since the seas are so intensively exploited and are surrounded by advanced industrialized countries, pollution problems are serious. The main pollution sources include rivers and other outfalls, dumping by ships (of dredged materials, sewage sludge and chemical wastes) and operational discharges from offshore installations and ships. Deposition of atmospheric pollutants is an additional major source of pollution.

Those parts of the seas at greatest risk from pollution are where the sediments come to rest, where the water replacement is slowest, and where nutrient concentrations and biological productivity are highest. A number of warning signals have occurred.

Algal populations have changed in number and species. There have been algal blooms, caused by excessive nutrient discharge from land and atmospheric sources. Species changes show a tendency toward more short-lived species of the opportunistic type and a reduction, sometimes to the point of disappearance, of some mammal, fish and sea grass species. Decreases of ray, mackerel, sand eel and echinoderms due to eutrophication have resulted in reduced plaice, cod, haddock and dab, mollusk and scoter. The impact of fishing activities is also considerable. Sea mammals, sea birds and Baltic fish species have been particularly affected by the widespread release of toxins and pollutants that accumulate in the sediments and in the food web. Some species, such as the grey seal and the sea eagle, are threatened with extinction.

Particular concern has been expressed about the Wadden Sea, which serves as a nursery for many North Sea species. Toxic PCB contamination, for example, almost caused the disappearance of seals in the 1970s.

The 1988 massive seal mortality in the North and Wadden Seas, although caused by a viral disease, is still thought by many to have a link with marine pollution.

Although the North Sea needs radical and lengthy treatment, it is probably not a terminal case. Actions are being taken by bordering countries to reduce the discharge of wastes into the sea. A major factor leading to agreements to reduce discharges of wastewaters has been the verification of predictive pollutant circulation models of the sea that identify the impacts of discharges from various sites along the sea boundary.

## 2.9. Egypt and the Nile: Limits to Agricultural Growth

Egypt, located in a belt of extreme aridity, is nearly completely dependent on the River Nile (Figure 1.15) for its water resources. Therefore, it is no wonder that most of Egypt's population lives close to the Nile. In relation to arable land and water, Egypt's population density is among the highest in the world: of its population of 63 million in 2000, 97% lives on 5% of land in the small strip along the Nile and in the Delta where water is abundant. The population density continues to increase as a result of a population growth of about 2% per year.

To relieve the population pressure in the Nile Delta and Nile Valley, the government has embarked on an ambitious programme to increase the inhabited area in Egypt from the present 5% to about 25% in the future. The agricultural area is to be enlarged by 'horizontal expansion', which should increase the agricultural area from 3.4 million ha in 1997 to 4.1 million in 2017. New industrial areas are planned in the desert, to be supplied by Nile water. Most of these new agricultural and industrial developments are based on public–private partnerships, requiring the government to give guarantees for the availability of water. The Toskha project in the south and the El-Salaam scheme in the Sinai are examples of this kind of development.

However, the availability of Nile water remains the same. Under the present agreement with Sudan, Egypt is allowed to use 55.5 billion $m^3$ of Nile water each year. That water is nearly completely used already and a further increase in demand will result in a lower availability of water per hectare. Additional measures can and will be taken to increase the efficiency of water use in Egypt, but that will not be sufficient. It is no wonder that Egypt is looking into possibilities to increase the supply by taking measures upstream in Sudan and Ethiopia. Examples are the construction of reservoirs on the Blue Nile in Ethiopia and the Jonglei Canal in Sudan that will partly drain the swamps in the Sudd and decrease the evaporation from them. Cooperation with the other (nine) countries in the Nile basin is essential to enable those developments (see Figure 1.15). Hence, Egypt is a strong supporter of the work of the Nile Basin Initiative that provides a framework for this cooperation. Other countries in the basin are challenging the claim of Egypt for additional water. If Egypt is unable to increase its supply, it will be forced to lower its ambitions on horizontal expansion of agriculture in the desert and to provide other means of livelihood for its growing population.

## 2.10. Damming the Mekong

The Mekong River (Figures 1.16 and 1.17) flows some 4,200 km through Southeast Asia to the South China Sea through Tibet, Myanmar (Burma), Vietnam, Laos, Thailand and Cambodia. Its 'development' has been restricted over the past several decades because of regional conflicts, indeed those that have altered the history of the world. Now that these conflicts are reduced, investment capital is becoming available to develop the Mekong's resources for improved fishing, irrigation, flood control, hydroelectric power, tourism, recreation and navigation. The potential benefits are substantial, but so are the environmental and ecological risks.

During some months of the year the lack of rainfall causes the Mekong to fall dramatically. Salt water may penetrate as much as 500 km inland. In other months the flow can be up to thirty times the low flows, causing the water in the river to back up into wetlands and flood some 12,000 $km^2$ of forests and paddy fields in the Vietnamese delta region alone. The ecology of a major lake, Tonle Sap in Cambodia, depends on these backed-up waters.

While flooding imposes risks on some 50 million inhabitants of the Mekong floodplain, there are also distinct advantages. High waters deposit nutrient-rich silts on the low-lying farmlands, thus sparing the farmers from having to transport and spread fertilizers on their fields. Also, shallow lakes and submerged lands provide spawning habitats for about 90% of the fish in the Mekong Basin. Fish yield totals over half a million tons annually.


**Figure 1.15.** The Nile Basin.

**Figure 1.16.** The Mekong River is one of the few rivers that is still in equilibrium with surrounding life.

What will happen to the social fabric and to the natural environment if the schemes to build big dams across the mainstream of the Mekong are implemented? Depending on their operation, they could disrupt the current fertility cycles, habitats and habits of the fish in the river. Increased erosion downstream from major reservoirs is also a threat. Add to these the possible adverse impacts the need to evacuate and resettle thousands of people displaced by the lake behind the dams. How will they be resettled? And how long will it take them to adjust to new farming conditions?

There have been suggestions that a proposed dam in Laos could cause deforestation in a wilderness area of some 3,000 $km^2$. Much of the wildlife, including elephants, big cats and other rare animals, would have to be protected if they are not to become endangered. Malaria-carrying mosquitoes, liver fluke and other disease bearers might find ideal breeding grounds in the mud flats of the shallow reservoir. These are the types of issues that need to be considered now that increased development seems possible, and even likely.

Consider, for example, the impacts of a dam constructed on the Nam Pong River in northeast Thailand. The Nam Pong project was to provide hydroelectric power and irrigation water, the avowed purposes of many reservoir projects throughout the world. Considerable attention was paid to the social aspects of this project, but not to the environmental impacts. The project had a number of unexpected consequences, both beneficial and adverse.

Because the reservoir was acting as a bioreactor for most of the year, the fish population became so large that a major fishery industry has developed around the reservoir. The economic benefits of fish production exceeded those derived from hydropower. However, lack of adequate planning for this development resulted in less than ideal living and economic conditions for the migrating fishermen who came to this region.

Despite the availability of irrigation water, most farmers were still practising single-crop agriculture after the dam was built, and still growing traditional crops in their traditional ways. No training was provided for them to adapt their skills to the new conditions and opportunities. In addition, while farming income did not decrease, the general welfare and health of the population seems to have deteriorated. Again, little attention was given to diet and hygiene under these new conditions.

The reservoir itself had some adverse impacts along with the beneficial ones. These included increased erosion of the stream banks, silting up of the channel and a large increase in aquatic vegetation that clogged hydraulic machinery and reduced transport capacity.

## 3. So, Why Plan, Why Manage?

Water resources planning and management activities are usually motivated, as they were in each of the previous section's case examples, by the realization that there are both problems to solve and opportunities to obtain increased benefits from the use of water and related land


**Figure 1.17.** The Lower Mekong River Basin.

resources. These benefits can be measured in many different ways. Inevitably, it is not easy to agree on the best way to do so, and whatever is proposed may provoke conflict. Hence there is the need for careful study and research, as well as full stakeholder involvement, in the search for a shared vision of the best compromised plan or management policy.

Reducing the frequency and/or severity of the adverse consequences of droughts, floods and excessive pollution are common goals of many planning and management exercises. Other goals include the identification and evaluation of alternative measures that may increase the available water supplies or hydropower, improve recreation and/or navigation, and enhance the quality of water and aquatic ecosystems. Quantitative system performance criteria can help one judge the relative net benefits, however measured, of alternative plans and management policies.

System performance criteria of interest have evolved over time. They have developed from being primarily focused on safe drinking water just a century ago, to multipurpose economic development a half-century ago, to goals that now include environmental and ecosystem restoration and protection, aesthetic and recreational experiences, and more recently, sustainability (ASCE, 1998).

Some of the multiple purposes served by a river can be conflicting. A reservoir used solely for hydropower or water supply is better able to meet its objectives when it is full of water, rather than when it is empty. On the other hand, a reservoir used solely for downstream flood control is best left empty, until the flood comes of course. A single reservoir serving all three purposes introduces conflicts over how much water to store in it and how it should be operated. In basins where diversion demands exceed the available supplies, conflicts will exist over water allocations. Finding the best way to manage, if not resolve, these conflicts that occur over time and space are other reasons for planning.

### 3.1. Too Little Water

Issues involving inadequate supplies to meet demands can result from conflicts or concerns over land and water use. They can result from growing urbanization, the development of additional water supplies, the need to meet instream flow requirements, and conflicts over private property and public rights regarding water allocations. Other issues can involve trans-basin water transfers and markets, objectives of economic efficiency versus the desire to keep non-efficient activities viable, and demand management measures, including incentives for water reuse and water reuse financing.

Measures to reduce the demand for water in times of supply scarcity should be identified and agreed upon before everyone has to cope with an actual water scarcity. The institutional authority to implement drought measures when their designated 'triggers' – such as decreasing storage volumes in reservoirs – have been met should be established before the measures are needed. Such management responses may include increased groundwater abstractions to supplement low surface-water flows and storage volumes. Conjunctive use of ground and surface waters can be sustainable as long as the groundwater aquifers are recharged during conditions of high flow and storage volumes.

### 3.2. Too Much Water

Damage due to flooding is a direct result of floodplain development that is vulnerable to floods. This is a risk many take, and indeed on average it may result in positive private net benefits, especially when public agencies subsidize these private risk takers in times of flooding. In many river basins of developed regions, the level of annual expected flood damage is increasing over time, in spite of increased expenditures on flood damage reduction measures. This is mainly due to increased economic development on river floodplains, not to increased frequencies or magnitudes of floods.

The increased economic value of the development on floodplains often justifies increased expenditure on flood damage reduction measures. Flood protection works decrease the risks of flooding and consequent damage, creating an incentive for increased economic development. Then when a flood exceeding the capacity of existing flood protection works occurs, and it will, even more damage results. This cycle of increasing flood damage and cost of protection is a natural result of the increasing values of floodplain development.

Just what is the appropriate level of risk? It may depend, as Figure 1.18 illustrates, on the level of flood insurance or subsidy provided when flooding occurs.

Flood damage will decrease only if restrictions are placed on floodplain development. Analyses carried out during planning can help identify the appropriate level of

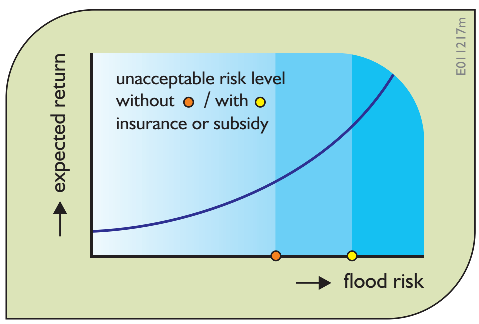



**Figure 1.18.** The lowest risk of flooding on a floodplain does not always mean the best risk, and what risk is acceptable may depend on the amount of insurance or subsidy provided when flood damage occurs.

development and flood damage protection works, on the basis of both the beneficial and the adverse economic, environmental and ecological consequences of floodplain development. People are increasingly recognizing the economic as well as environmental and ecological benefits of allowing floodplains to do what they were formed to do: store flood waters when floods occur.

## 3.3. Polluted Water

The discharges of wastewater by industry and households can have considerable detrimental effects on water quality and, hence, often on public and ecosystem health. Planning and management activities should pay attention to these possible negative consequences of industrial development, population growth and the intensive use of pesticides and fertilizers in urban as well as in agricultural areas. Issues regarding the environment and water quality include:

- upstream versus downstream conflicts on meeting water quality standards
- threats from aquatic nuisance species
- threats from the chemical, physical and biological water quality of the watershed's aquatic resources
- quality standards for recycled water
- non-point source pollution discharges, including sediment from erosion
- inadequate groundwater protection compacts and concerned institutions.

We still know too little about the environmental and health impacts of many of the wastewater constituents found in river waters. As more is learned about, for example, the harmful effects of heavy metals and dioxins, our plans and management policies should be adjusted accordingly. Major fish losses and algae blooms point to the need to manage water quality as well as quantity.

## 3.4. Degradation of Aquatic and Riparian Ecosystems

Aquatic and riparian ecosystems may be subject to a number of threats. The most important include habitat loss due to river training and reclamation of floodplains and wetlands for urban and industrial development, poor water quality due to discharges of pesticides, fertilizers and wastewater effluents, and the infestation of aquatic nuisance species.

Exotic aquatic nuisance species can be major threats to the chemical, physical and biological water quality of a river's aquatic resources, and a major interference with other uses. The destruction and/or loss of the biological integrity of aquatic habitats caused by introduced exotic species is considered by many ecologists to be among the most important problems facing natural aquatic and terrestrial ecosystems. The biological integrity of natural ecosystems is controlled by habitat quality, water flows or discharges, water quality and biological interactions including those involving exotic species.

Once exotic species are established, they are usually difficult to manage and nearly impossible to eliminate. This creates a costly burden for current and future generations. The invasion in North America of non-indigenous aquatic nuisance species such as the sea lamprey, zebra mussel, purple loosestrife, European green crab and various aquatic plant species, for example, has had pronounced economic and ecological consequences for all who use or otherwise benefit from aquatic ecosystems.

Environmental and ecological effectiveness as well as economic efficiency should be a guiding principle in evaluating alternative solutions to problems caused by aquatic nuisance organisms. Funds spent on proper prevention and early detection and eradication of aquatic nuisance species may reduce the need to spend considerably greater funds on management and control once such species are well established.

## 3.5. Other Planning and Management Issues

### Navigation

Industrial and related port development may result in the demand for deeper rivers to allow the operation of larger-draught cargo vessels in the river. River channel improvement cannot be detached from functions such as water supply and flood control. Narrowing the river for shipping purposes may increase floodwater levels.

### River Bank Erosion

Bank erosion can be a serious problem where people are living close to morphologically active (eroding) rivers.

Bangladesh, where bank erosion is considered to be a much more urgent problem than the well-known floods of that country, is an example of this. Predictions of changes in river courses due to bank erosion and bank accretion are important inputs to land use planning in river valleys and the choice of locations for bridges and hydraulic structures.

### Reservoir Related Issues

Degradation of the riverbed upstream of reservoirs may increase the risks of flooding in those areas. Reservoir construction inevitably results in loss of land and forces the evacuation of residents due to impoundment. Dams can be ecological barriers for migrating fish species such as salmon. The water and sediment quality in the reservoir may deteriorate and the in-flowing sediment may accumulate, reducing the active (useful) capacity of the reservoir. Other potential problems may include those stemming from stratification, water related diseases, algae growth and abrasion of hydropower turbines.

Environmental and morphological impacts downstream of the dam are often due to a changed river hydrograph and decreased sediment load in the water released from the reservoir. Lower sediment loads result in higher scouring of downstream riverbeds and consequently a lowering of their elevations. Economic as well as social impacts include the risk of dams breaking. Environmental impacts may result from sedimentation control measures (e.g., sediment flushing) and reduced oxygen content of the out-flowing water.

The ecological, environmental and economic impacts of dams and reservoirs are heavily debated among planners and environmentalists. In creating a new framework for decision-making, the World Commission on Dams compiled and considered the arguments of all sides of this debate (WCD, 2000).

## 4. System Components, Planning Scales and Sustainability

Water resources management involves influencing and improving the interaction of three interdependent subsystems:

- the natural river subsystem in which the physical, chemical and biological processes take place
- the socio-economic subsystem, which includes the human activities related to the use of the natural river system
- the administrative and institutional subsystem of administration, legislation and regulation, where the decision and planning and management processes take place.

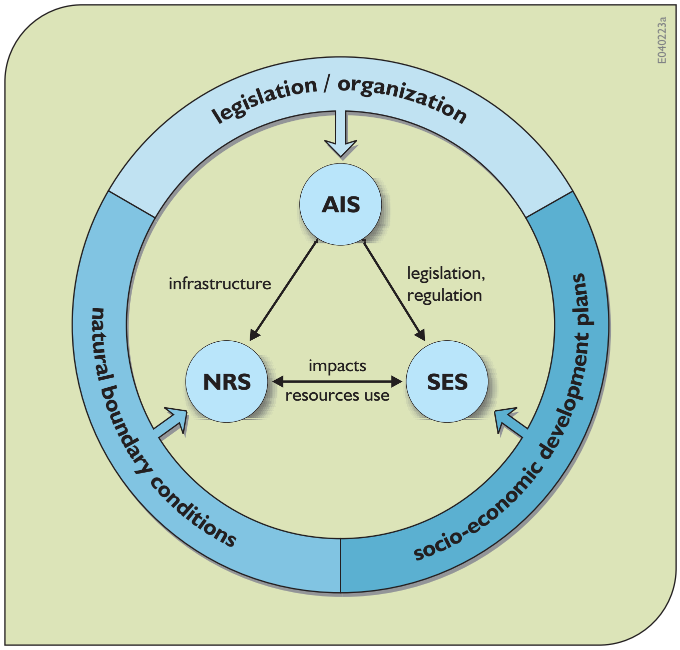



Figure 1.19. Interactions among subsystems and between them and their environment.

Figure 1.19 illustrates the interaction between these subsystems, all three of which should be included in any analysis performed for water resource systems planning and management. Inadequate attention to one can destroy the value of any work done to improve the performance of the others.

Appendix A describes the major components of the natural system and their processes and interactions.

### 4.1. Spatial Scales for Planning and Management

Watersheds or river basins are usually considered logical regions for water resources planning and management. This makes sense if the impacts of decisions regarding water resources management are contained within the watershed or basin. How land and water are managed in one part of a river basin can affect the land and water in other parts of the basin. For example, the discharge of pollutants or the clearing of forests in the upstream portion of the basin may degrade the quality and increase the variability of the flows and sedimentation downstream. The construction of a dam or weir in the downstream part of a river may prevent vessels and fish from travelling upstream. To maximize the economic and social benefits obtained from the entire basin, and to ensure that these benefits and accompanying costs are equitably distributed, planning and management is often undertaken on a basin scale.

While basin boundaries make sense from a hydrological point of view, they may be inadequate for addressing particular water resources problems that are caused by events taking place outside the basin. What is desired is the highest level of performance, however defined, of the entire physical, socio-economic and administrative water resource system. To the extent that the applicable problems, stakeholders and administrative boundaries extend outside the river basin, the physically based 'river basin' focus of planning and management should be expanded to include the entire applicable 'problem-shed'. Hence, consider the term 'river basin' used in this book to mean problem-shed when appropriate.

## 4.2. Temporal Scales for Planning and Management

Water resources planning requires looking into the future. Decisions recommended for the immediate future should take account of their long-term future impacts. These impacts may also depend on economic, demographic and physical conditions now and on into some distant future. The question of just how far into the future one need look, and try to forecast, is directly dependent on the influence that future forecast has on the present decisions. What is most important now is what decision to make now. Decisions that are to be made later can be based on updated forecasts, then-current information and planning and management objectives. Planning is a continuing sequential process. Water resources plans need to be periodically updated and adapted to new information, new objectives, and updated forecasts of future supplies, demands, costs and benefits.

The number and duration of within-year time periods explicitly considered in the planning process will be dependent in part on the need to consider the variability of the supplies and demands for water resources and on the purposes to be served by the water resources within the basin. Irrigation planning and summer-season water recreation planning may require a greater number of within-year periods during the summer growing and recreation season than might be the case if one were considering only municipal water supply planning, for example. Assessing the impacts of alternatives for conjunctive surface and groundwater management, or for water quantity and quality management, require attention to processes that take place on different spatial and temporal scales.

## 4.3. Sustainability

Sustainable water resources systems are those designed and managed to best serve people living today and in the future. The actions that we as a society take now to satisfy our own needs and desires should depend not only on what those actions will do for us but also on how they will affect our descendants. This consideration of the long-term impacts on future generations of actions taken now is the essence of sustainable development. While the word 'sustainability' can mean different things to different people, it always includes a consideration of the welfare of those living in the future. While the debate over a more precise definition of sustainability will continue, and questions over just what it is that should be sustained may remain unanswered, this should not delay progress toward achieving more sustainable water resources systems.

The concept of environmental and ecological sustainability has largely resulted from a growing concern about the long-run health of our planet. There is increasing evidence that our present resource use and management activities and actions, even at local levels, can significantly affect the welfare of those living within much larger regions in the future. Water resource management problems at a river basin level are rarely purely technical and of interest only to those living within the individual river basins where those problems exist. They are increasingly related to broader societal structures, demands and goals.

What would future generations like us to do for them? We don't know, but we can guess. As uncertain as these guesses will be, we should take them into account as we act to satisfy our own immediate needs, demands and desires. There may be tradeoffs between what we wish to do for ourselves in our current generation versus what we think future generations might wish us to do for them. These tradeoffs, if any, between what present and future generations would like should be considered and debated in the political arena. There is no scientific theory to help us identify which tradeoffs, if any, are optimal.

The inclusion of sustainability criteria along with the more common economic, environmental, ecological and social criteria used to evaluate alternative water resources development and management strategies may identify a need to change how we commonly develop and use our water resources. We need to consider the impacts of change itself. Change over time is certain – just what it will be is uncertain. These changes will affect the physical, biological and social dimensions of water resource systems. An essential aspect in the planning, design and management of sustainable systems is the anticipation of change. This includes change due to geomorphologic processes, the aging of infrastructure, shifts in demands or desires of a changing society, and even increased variability of water supplies, possibly because of a changing climate. Change is an essential feature of sustainable water resources development and management.

Sustainable water resources systems are those designed and operated in ways that make them more adaptive, robust and resilient to an uncertain and changing future. They must be capable of functioning effectively under conditions of changing supplies, management objectives and demands. Sustainable systems, like any others, may fail, but when they fail they must be capable of recovering and operating properly without undue costs.

In the face of certain changes, but with uncertain impacts, an evolving and adaptive strategy for water resources development, management and use is a necessary condition of sustainable development. Conversely, inflexibility in the face of new information, objectives and social and political environments is an indication of reduced system sustainability. Adaptive management is a process of adjusting management actions and directions, as appropriate, in the light of new information on the current and likely future condition of our total environment and on our progress toward meeting our goals and objectives. Water resources development and management decisions can be viewed as experiments, subject to modification, but with goals clearly in mind. Adaptive management recognizes the limitations of current knowledge and experience as well as those that we learn by experimenting. It helps us move toward meeting our changing goals over time in the face of this incomplete knowledge and uncertainty. It accepts the fact that there is a continual need to review and revise management approaches because of the changing, as well as uncertain, nature of our socio-economic and natural environments.

Changing the social and institutional components of water resources systems is often the most challenging task, because it involves changing the way individuals think and act. Any process involving change will require that we change our institutions – the rules under which we as a society function. Individuals are primarily responsible for, and adaptive to, changing political and social situations. Sustainability requires that public and private institutions also change over time in ways that are responsive to the needs of individuals and society.

Given the uncertainty of what future generations will want, and the economic, environmental and ecological problems they will face, a guiding principle for the achievement of sustainable water resource systems is to provide options to future generations. One of the best ways to do this is to interfere as little as possible with the proper functioning of natural life cycles within river basins, estuaries and coastal zones. Throughout the water resources system planning and management process, it is important to identify all the beneficial and adverse ecological, economic, environmental and social effects – especially the long-term effects – associated with any proposed project.

# 5. Planning and Management

## 5.1. Approaches

There are two general approaches to planning and management. One is from the top down, often called command and control. The other is from the bottom up, often called a grass-roots approach. Both approaches can lead to an integrated plan and management policy.

### 5.1.1. Top-Down Planning and Management

Over much of the past half century, water resources professionals have been engaged in preparing integrated, multipurpose 'master' development plans for many of the world's river basins. These plans typically consist of a series of reports, complete with numerous appendices, describing all aspects of water resources management and use. In these documents alternative structural and non-structural management options are identified and evaluated. On the basis of these evaluations, the preferred plan is presented.

This master planning exercise has typically been a top-down approach that professionals have dominated. Using this approach there is usually little if any active participation by interested stakeholders. The approach assumes that one or more institutions have the ability and authority to develop and implement the plan, in other words, that will oversee and manage the coordinated development and operation of the basin's activities that affect the surface and ground waters of the basin. In today's environment, where publics are calling for less governmental oversight, regulation and control, and increasing participation in planning and management activities, top-down approaches are becoming less desirable or acceptable.

### 5.1.2. Bottom-Up Planning and Management

Within the past decade water resources planning and management processes have increasingly involved the active participation of interested stakeholders – those affected in any way by the management of the water and land resources. Plans are being created from the bottom up rather than top down. Concerned citizens and non-governmental organizations, as well as professionals in governmental agencies, are increasingly working together towards the creation of adaptive comprehensive water management programs, policies and plans.

Experiences of trying to implement plans developed primarily by professionals without significant citizen involvement have shown that, even if such plans are technically flawless, they have little chance of success if they do not take into consideration the concerns of affected local stakeholders and do not have their support. To gain this, concerned stakeholders must be included in the decision-making process as early as possible. They must become part of that process, not merely as spectators or advisors to it. This will help gain their cooperation and commitment to the plans adopted. Participating stakeholders will have a sense of ownership, and as such will strive to make the plans work. Such plans, if they are to be successfully implemented, must also fit within existing legislative, permitting, enforcement and monitoring programmes. Stakeholder participation improves the chance that the system being managed will be sustainable.

Successful planning and management involves motivating all potential stakeholders and sponsors to join in the water resources planning and management process, determining their respective roles and establishing how to achieve consensus on goals and objectives. Ideally this should occur before addressing conflicting issues so that all involved know each other and are able to work together more effectively. Agreements on goals and objectives and on the organization (or group formed from multiple organizations) that will lead and coordinate the water resources planning and management process should be reached before stakeholders bring their individual priorities or problems to the table. Once the inevitable conflicts become identified, the settling of administrative matters doesn't get any easier.

Bottom-up planning must strive to achieve a common or 'shared' vision of goals and priorities among all stakeholders. It must be aware of and comply with all applicable laws and regulations. It should strive to identify and evaluate multiple alternatives and performance criteria – including sustainability criteria – and yet keep the process from producing a wish-list of everything each stakeholder wants. In other words, it must identify tradeoffs among conflicting goals or measures of performance, and prioritize appropriate strategies. It must value and compare, somehow, the intangible and non-monetary impacts of environmental and ecosystem protection and restoration with other activities whose benefits and costs can be expressed in monetary units. In doing so, planners should use modern information technology to improve both the process and product. This technology, however, will not eliminate the need to reach conclusions and make decisions on the basis of incomplete and uncertain data and scientific knowledge.

These process issues focus on the need to make water resources planning and management as efficient and effective as possible. Many issues will arise in terms of evaluating alternatives and establishing performance criteria (prioritizing issues and possible actions), performing incremental cost analysis, and valuing monetary and non-monetary benefits. Questions must be answered as to how much data must be collected and with what precision, and what types of modern information technology (e.g., geographic information systems (GIS), remote sensing, Internet, decision support systems, etc.) can be beneficially used for both analyses and communication.

### 5.1.3. Integrated Water Resources Management

The concept of integrated water resources management (IWRM) has been developing since the beginning of the eighties. IWRM is the response to the growing pressure on our water resources systems caused by growing population and socio-economic developments. Water shortages and deteriorating water quality have forced many countries in the world to reconsider their options with respect to the management of their water resources. As a result water resources management (WRM) has been undergoing a change worldwide, moving from a mainly supply-oriented, engineering-biased approach towards a demand-oriented, multi-sectoral approach, often labelled integrated water resources management.

In international meetings, opinions are converging to a consensus about the implications of IWRM. This is best reflected in the Dublin Principles of 1992 (see Box 1.1), which have been universally accepted as the base for IWRM. The concept of IWRM makes us move away from top-down 'water master planning' (see Section 5.1.1), which focuses on water availability and development, towards 'comprehensive water policy planning' which addresses the interaction between different sub-sectors, seeks to establish priorities, considers institutional requirements and deals with the building of management capacity.

IWRM considers the use of the resources in relation to social and economic activities and functions. These also determine the need for laws and regulations for the sustainable use of the water resources. Infrastructure made available, in relation to regulatory measures and mechanisms, will allow for effective use of the resource, taking due account of the environmental carrying capacity (Box 1.2).

**Box 1.1. The Dublin Principles**

1. Water is a finite, vulnerable and essential resource, essential to sustain life, development and the environment.
2. Water resources development and management should be based on a participatory approach, involving users, planners and policy makers at all levels.
3. Women play a central role in the provision, management and safeguarding of water.
4. Water has an economic value in all its competing uses and should be recognized as an economic good.

**Box 1.2. Definition of IWRM**

IWRM is a *process* which promotes the coordinated development and management of water, land and related resources, in order to maximize the resultant *economic and social welfare* in an equitable manner without compromising the *sustainability of vital ecosystems*.

(GWP, 2000)

## 5.2. Planning and Management Aspects

### 5.2.1. Technical Aspects

Technical aspects of planning include hydrological assessments. These identify and characterize the properties of, and interactions among, the resources in the basin or region, including the land, the rainfall, the runoff, the stream and river flows and the groundwater.

Existing watershed land use and land cover, and future changes in this use and cover, result in part from existing and future changes in regional population and economy. Planning involves predicting changes in land use/covers and economic activities at watershed and river basin levels. These will influence the amount of runoff, and the concentrations of sediment and other quality constituents (organic wastes, nutrients, pesticides, etc.) it contains as a result of any given pattern of rainfall over the land area. These predictions will help planners estimate the quantities and qualities of flows and their constituents throughout a watershed or basin, associated with any land use and water management policy. This in turn provides the basis for predicting the type and health of terrestrial and aquatic ecosystems in the basin. All of this may affect the economic development of the region, which in part determines the future demands for changes in land use and land cover.

Technical aspects also include the estimation of the costs and benefits of any measures taken to manage the basin's water resources. These measures might include:

- engineering structures for making better use of scarce water
- canals and water-lifting devices
- dams and storage reservoirs that can retain excess water from periods of high-flow for use during the periods of low-flow (and may reduce flood damage below the reservoir by storing floodwater)
- open channels that may take the form of a canal, flume, tunnel or partly filled pipe
- pressure conduits
- diversion structures, ditches, pipes, checks, flow dividers and other engineering facilities necessary for the effective operation of irrigation and drainage systems
- municipal and industrial water intakes, including water purification plants and transmission facilities
- sewerage and industrial wastewater treatment plants, including waste collection and ultimate disposal facilities
- hydroelectric power storage, run-of-river or pumped storage plants,
- river channel regulation works, bank stabilization, navigation dams and barrages, navigation locks and other engineering facilities for improving a river for navigation
- levees and floodwalls for confinement of the flow within a predetermined channel.

Not only must the planning process identify and evaluate alternative management strategies involving structural and non-structural measures that will incur costs and bring benefits, but it must also identify and evaluate alternative time schedules for implementing those measures. The planning of development over time involving interdependent projects, uncertain future supplies and demands as well as costs, benefits and interest (discount) rates is part of all water resources planning and management processes.

With increasing emphasis placed on ecosystem preservation and enhancement, planning must include ecologic impact assessments. The mix of soil types and depths and land covers together with the hydrological quantity and quality flow and storage regimes in rivers, lakes, wetlands and aquifers affect the riparian and aquatic ecology of the basin. Water managers are being asked to consider ways of improving or restoring ecosystems by, for example, reducing:

- the destruction and/or loss of the biological integrity of aquatic habitats caused by introduced exotic species
- the decline in number and extent of wetlands and the adverse impacts on wetlands of proposed land and water development projects
- the conflicts between the needs of people for water supply, recreation, energy, flood control, and navigation infrastructure and the needs of ecological communities, including endangered species.

And indeed there are and will continue to be conflicts among alternative objectives and purposes of water management. Planners and managers must identify the trade-offs among environmental, ecologic, economic and social impacts, however measured, and the management alternatives that can balance these often-conflicting interests.

### 5.2.2. Economic and Financial Aspects

The fourth Dublin principle states that water has an economic value in all its competing uses and should be recognized as an economic good. This principle addresses the need to extract the maximum benefits from a limited resource as well as the need to generate funds to recover the costs of the investments and of the operation and maintenance of the system.

The maximization of benefits is based on a common economic market approach. Many past failures in water resources management are attributable to the fact that water has been – and still is – viewed as a free good. Prices of water for irrigation and drinking water are in many countries well below the full cost of the infrastructure and personnel needed to provide that water, which comprises the capital charges involved, the operation and maintenance (O&M) costs, the opportunity cost, economic externalities and environmental externalities (see GWP, 2000). Charging for water at less than full cost means that the government, society and/or environment 'subsidizes' water use and leads to sub-optimal use of the resource.

Recognizing water as an economic good does not always mean that full costs should be charged. Poor people have the right to safe water and this should be taken into account. For that reason the fourth Dublin principle is often referred to as water being an economic and social good.

Cost recovery is the second reason for the fourth Dublin principle. The overriding financial component of any planning process is to make sure that the recommended plans and projects are able to pay for themselves. Revenues are needed to recover construction costs, if any, and to maintain, repair and operate any infrastructure designed to manage the basin's water resources. This may require cost-recovery policies that involve pricing the outputs of projects. Beneficiaries should be expected to pay at least something, and in some way, for the added benefits they get. Planning must identify equitable cost and risk-sharing policies and improved approaches to risk/cost management. In many developing countries a distinction is made between cost recovery of investments and cost recovery of O&M costs. Cost recovery of O&M costs is a minimum condition for a sustainable project. Without that, it is likely that the performance of the project will deteriorate seriously over time.

In most WRM studies, financial viability is viewed as a constraint that must be satisfied. It is not viewed as an objective whose maximization could result in a reduction in economic efficiency, equity or other non-monetary objectives.

### 5.2.3. Institutional Aspects

The first condition for successful project implementation is to have an enabling environment. There must exist national, provincial and local policies, legislation and institutions that make it possible for the right decisions to be taken and implemented correctly. The role of the government is crucial. The reasons for governmental involvement are manifold:

- Water is a resource beyond property rights: it cannot be 'owned' by private persons. Water rights can be given to persons or companies, but only the rights to use the water and not to own it. Conflicts between users automatically turn up at the table of the final owner of the resource – the government.
- Water is a resource that often requires large investment to develop. Many water resources development projects are very expensive and have many beneficiaries. Examples are multipurpose reservoirs and the construction of dykes along coasts and rivers. The required investments need large financial commitments which only can be made by the government or state-owned companies.
- Water is a medium that can easily transfer external effects. The use of water by one person often has negative effects on others (externalities). The obvious example is the discharge of waste into a river that may have negative effects on downstream users.

Only the government can address these issues and 'good governance' is necessary for good water management.

An insufficient institutional setting and the lack of a sound economic base are the main causes of water resources development project failure, not technical inadequacy of design and construction. This is also the reason why at present much attention is given to institutional developments in the sector, in both developed and developing countries. In Europe, various types of water agencies are operational (e.g., the Agence de l'Eau in France and the water companies in England), each having advantages and disadvantages. The Water Framework Directive of the European Union requires that water management be carried out at the scale of a river basin, particularly when this involves transboundary management. It is very likely that this will result in a shift in responsibilities of the institutions involved and the establishment of new institutions. In other parts of the world experiments are being carried out with various types of river basin organizations, combining local, regional and sometimes national governments.

## 5.3. Analyses for Planning and Management

Analyses for water resources planning and management generally comprise several stages. The explicit description of these stages is referred to as the *analytical* (or conceptual) framework. Within this framework, a set of coherent models for the quantitative analysis of measures and strategies is used. This set of models and related databases will be referred to as the *computational* framework. This book is mainly about the computational framework.

The purpose of the analyses is to prepare and support planning and management decisions. The main phases of the analytical framework therefore correspond to the phases of the decision process. Such a decision process is not a simple, one-line sequence of steps. Inherent in a decision-making process are factors causing the decision-makers to return to earlier steps of the process. Part of the process is thus cyclic. A distinction is made between comprehension cycles and feedback cycles. A *comprehension* cycle improves the decision-makers' understanding of a complex problem by cycling within or between steps. *Feedback* cycles imply returning to earlier phases of the process. They are needed when:

- solutions fail to meet criteria.
- new insights change the perception of the problem and its solutions (e.g., due to more/better information).
- essential system links have been overlooked.
- situations change (political, international, societal developments).

As an example, the analytical framework that is used by Delft Hydraulics for WRM studies is depicted in Figure 1.20. The three elementary phases of that framework are:

- inception
- development
- selection.

During each phase the processes have a cyclic component (comprehensive cycle). Interaction with the decision-makers, or their representatives, is essential throughout the process. Regular reporting through inception and interim reports will improve the effectiveness of the communication.

The first phase of the process is the inception phase. Here the subject of the analysis (what is analysed under

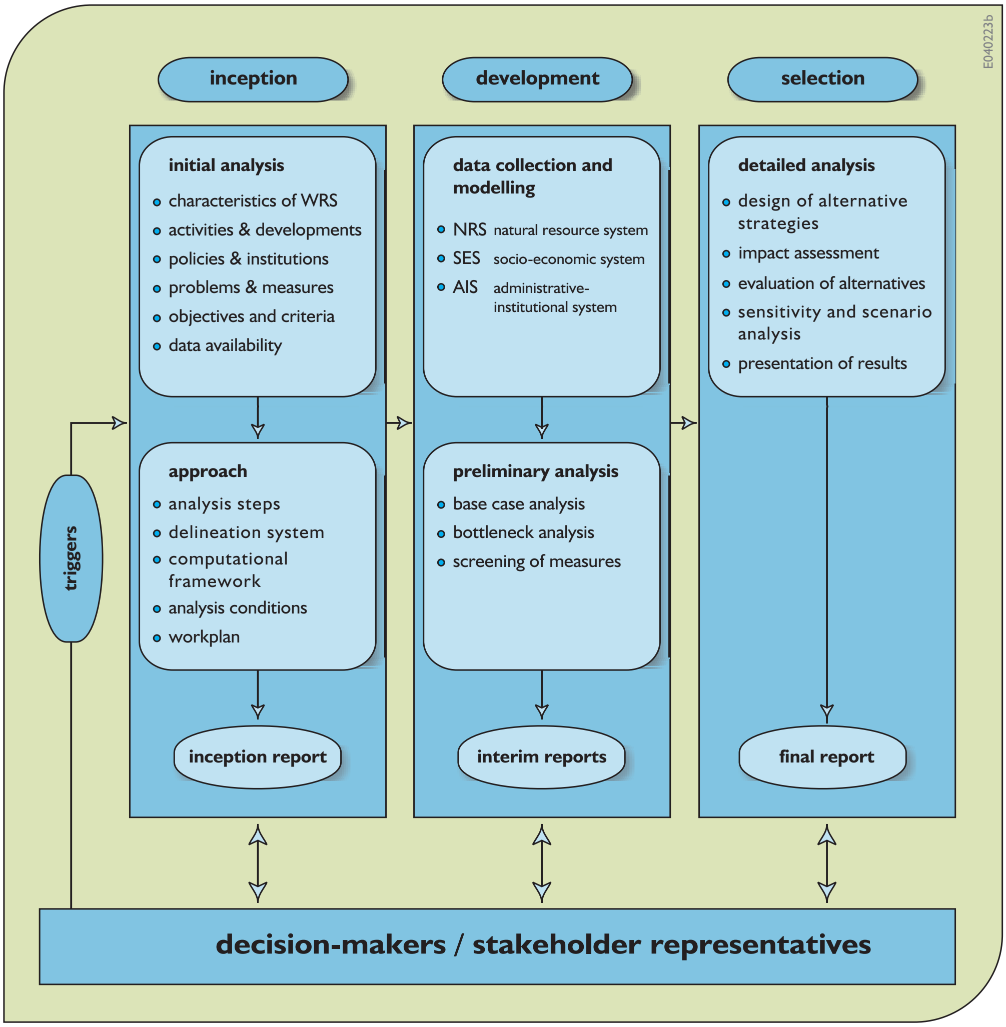



**Figure 1.20.** Typical analytical framework for water resources studies.

what conditions) and its object (the desired results of the analysis) are specified. Based on this initial analysis, during which intensive communication with (representatives of the) decision-makers is essential, the approach for the policy analysis is specified. The results of the inception phase are presented in the inception report, which includes the work plan for the other phases of the analysis process (project).

In the development phase tools are developed for analysing and identifying possible solutions to the WRM problems. The main block of activities is usually related to data collection and modelling. Various preliminary analyses will be made to ensure that the tools developed for the purpose are appropriate for solving the WRM problems. Individual measures will be developed and screened in this phase, and preliminary attempts will be made to combine promising measures into management strategies. The development phase is characterized by an increased understanding of the functioning of the water resources system, starting with limited data sets and simplified tools and ending at the levels of detail deemed necessary in the inception phase. Scanning of possible measures should also start as soon as possible during this phase. The desired level of detail in the data collection and modelling strongly depends on what is required to distinguish among the various measures being considered. Interactions with decision-makers are facilitated through the presentation of interim results in interim reports.

The purpose of the selection phase is to prepare a limited number of promising strategies based on a detailed analysis of their effects on the evaluation criteria, and to present them to the decision-makers, who will make the final selection. Important activities in this phase are strategy design, evaluation of strategies and presentation. The results of this phase are included in the final report

Although it is clear that analyses are made to support the decision-making process, it is not always clear who will make the final decision, or who is the decision-maker. If analyses are contracted to a consultant, careful selection of the appropriate coordinating agency is instrumental to the successful implementation of the project. It is always advantageous to use existing line agencies as much as possible. Interactions with the decision-makers usually take place through steering commissions (with an interdepartmental forum) and technical advisory committees. Appendix E of this book describes this analytical framework in more detail.

## 5.4. Models for Impact Prediction and Evaluation

The process of planning has undergone a significant transformation over the past several decades, mainly due to the continuing development of improved computational technology, and various water resource simulation and optimization models together with their associated databases and user-friendly interactive interfaces. Planning today is heavily dependent on the use of computer-based impact prediction models. Such models are used to assist in the identification and evaluation of alternative ways of meeting various planning and management objectives. They provide an efficient way of analysing spatial and temporal data in an effort to predict the interaction and impacts, over space and time, of various river basin components under alternative designs and operating policies.

Many of the systems analysis approaches and models discussed in the accompanying chapters of this book have been, and continue to be, central to the planning and management process. Their usefulness is directly dependent on the quality of the data and models being used. Models can assist planning and management at different levels of detail. Some are used for preliminary screening of alternative plans and policies, and as such do not require major data collection efforts. Screening models can also be used to estimate how significant certain data and assumptions are for the decisions being considered, and hence can help guide additional data collection activities. At the other end of the planning and management spectrum, much more detailed models can be used for engineering design. These more complex models are more data demanding, and typically require higher levels of expertise for their proper use.

The integration of modelling technology into the social and political components of the planning and management processes in a way that enhances those processes continues to be the main challenge of those who develop planning and management models. Efforts to build and apply interactive generic modelling programs or 'shells,' on which interested stakeholders can 'draw in' their system, enter their data and operating rules at the level of detail desired, run simulations, and discover the effect of alternative assumptions and operating rules, has in many cases helped to create a common or shared understanding among these stakeholders. Getting stakeholders involved in developing and experimenting with their own interactive data-driven models has been an effective way of building a consensus–a shared vision.

### 5.5. Shared-Vision Modelling

Participatory planning inevitably involves conflict management. Each stakeholder or interest group has its objectives, interests and agendas. Some of these may be in conflict with others. The planning and management process is one of negotiation and compromise. This takes time but, from it can come decisions that have the best chance of being considered right and fair or equitable by most participants. Models can assist in this process of reaching a common understanding and agreement among different stakeholders. This has a greater chance of happening if the stakeholders themselves are involved in the modelling and analysis process.

Involving stakeholders in model-building accomplishes a number of things. It gives them a feeling of ownership. They will have a much better understanding of just what their model can do and what it cannot. If they are involved in model-building, they will know the assumptions built into their model.

Being involved in a joint modelling exercise is a way to understand better the impacts of various assumptions. While there may be no agreement on the best of various assumptions to make, stakeholders can learn which of those assumptions matter and which do not. In addition, the process of model development by numerous stakeholders will itself create discussions that will lead toward a better understanding of everyone's interests and concerns. Though such a model-building exercise, it is possible those involved will reach not only a better understanding of everyone's concerns, but also a common or 'shared' vision of at least how their system (as represented by their model, of course) works.

### 5.6. Adaptive Integrated Policies

One of the first issues to address when considering water resources planning and management activities is the product desired. If it is to be a report, what should the report contain? If it is to be a model or a decision support system, what should be its capabilities?

Clearly a portion of any report should contain a discussion of the water resources management issues and options. Another part of the report might include a prioritized list of strategies for addressing existing problems and available development or management opportunities in the basin.

Recent emphasis has shifted from structural engineering solutions to more non-structural alternatives, especially for environmental and ecosystem restoration. Part of this shift reflects the desire to keep more options open for future generations. It reflects the desire to be adaptive to new information and to respond to surprises – impacts not forecasted. As we learn more about how river basins, estuaries and coastal zones work, and how humans can better manage those resources, we do not want to regret what we have done in the past that may preclude this adaptation.

In some situations it may be desirable to create a 'rolling' plan – one that can be updated at any time. This permits responses to resource management and regulatory questions when they are asked, not just at times when new planning and management exercises take place. While this appears to be desirable, will planning and management organizations have the financing and support to maintain and update the modelling software used to estimate various impacts, collect and analyse new data, and maintain the expertise, all of which are necessary for continuous planning (rolling plans)?

Consideration also needs to be given to improving the quality of the water resources planning and management review process, and focusing on outcomes themselves rather than output measures. One of the outcomes should be an increased understanding of some of the relationships between various human activities and the hydrology and ecology of the basin, estuary or coastal zone. Models developed for predicting the economic as well as ecologic interactions and impacts due to changes in land and water management and use could be used to address questions such as:

- What are the hydrological, ecological and economic consequences of clustering or dispersing human land uses such as urban and commercial developments and large residential areas? Similarly, what are the consequences of concentrated versus dispersed patterns of reserve lands, stream buffers and forestland?

- What are the costs and ecological benefits of a conservation strategy based on near-stream measures (e.g., riparian buffers) versus near-source (e.g., upland/site-edge) measures? What is the relative cost of forgone upland development versus forgone valley or riparian development? Do costs strongly limit the use of stream buffer zones for mitigating agriculture, residential and urban developments?
- Should large intensive developments be best located in upland or valley areas? Does the answer differ depending on economic, environmental or aquatic ecosystem perspectives? From the same perspectives, is the most efficient and desirable landscape highly fragmented or highly zoned with centres of economic activity?
- To what extent can riparian conservation and enhancement mitigate upland human land use effects? How do the costs of upland controls compare with the costs of riparian mitigation measures?
- What are the economic and environmental quality tradeoffs associated with different areas of various classes of land use such as commercial/urban, residential, agriculture and forest?
- Can adverse effects on hydrology, aquatic ecology and water quality of urban areas be better mitigated through upstream or downstream management approaches? Can land controls like stream buffers be used at reasonable cost within urban areas, and if so, how effective are they?
- Is there a threshold size for residential/commercial areas that yield marked ecological effects?
- What are the ecological states at the landscape scale that, once attained, become irreversible with reasonable mitigation measures? For example, once stream segments in an urban setting become highly altered by direct and indirect effects (e.g., channel bank protection and straightening and urban runoff), can they be restored with feasible changes in urban land use or mitigation measures?
- Mitigating flood risk by minimizing floodplain developments coincides with conservation of aquatic life in streams. What are the economic costs of this type of risk avoidance?
- What are the economic limitations and ecological benefits of having light residential zones between waterways and commercial, urban or agricultural lands?
- What are the economic development decisions that are irreversible on the landscape? For example, once land is used for commercial development, it is normally too costly to return it to agriculture. This would identify limits on planning and management for conservation and development.
- What are the associated ecological and economic impacts of the trend in residential, commercial and forest lands replacing agricultural lands?

The answers to these and similar questions may well differ in different regions. However, if we can address them on a regional scale – in multiple river basins – we might begin to understand and predict better the interactions among economy, environment and ecology as a function of how we manage and use its land and water. This in turn may help us better manage and use our land and water resources for the betterment of all – now and in the future.

### 5.7. Post-Planning and Management Issues

Once a plan or strategy is produced, common implementation issues include seeing that the plan is followed, and modified, as appropriate, over time. What incentives need to be created to ensure compliance? How are the impacts resulting from the implementation of any decision going to be monitored, assessed and modified as required and desired? Who is going to be responsible? Who is going to pay, and how much? Who will keep the stakeholders informed? Who will keep the plan current? How often should plans and their databases be updated? How can new projects be operated in ways that increase the efficiencies and effectiveness of joint operation of multiple projects in watersheds or river basins – rather than each project being operated independently of the others? These questions should be asked and answered, at least in general terms, before the water resources planning and management process begins. The questions should be revisited as decisions are made and when answers to them can be much more specific.

## 6. Meeting the Planning and Management Challenges: A Summary

Planning (the formulation of development and management plans and policies) is an important and often indispensable means to support and improve operational management. It provides an opportunity to:

- assess the current state of the water resources and the conflicts and priorities over their use, formulate visions, set goals and targets, and thus orient operational management
- provide a framework for organizing policy relevant research and public participation
- increase the legitimacy, public acceptance of (or even support for) the way the resources are to be allocated or controlled, especially in times of stress
- facilitate the interaction, discussion and coordination among managers and stakeholders, and generate a common point of reference – a management plan or policy.

Many of the concerns and issues being addressed by water resources planners and managers today are similar to those faced by planners and managers in the past. But some are different. Most of the new ones are the result of two trends: first, a growing concern for the sustainability of natural ecosystems and second, an increased recognition of the need of a bottom-up 'grass-roots' participatory approach to planning, managing and decision-making.

Today planners work for economic development and prosperity as they did in the past, keeping in mind environmental impacts and goals as they did then, but now recognizing ecological impacts and values as well. Water resources management may still be focused on controlling and mitigating the adverse impacts of floods and droughts and water pollution, on producing hydropower, on developing irrigation, on controlling erosion and sediment, and on promoting navigation, but only as these and similar activities are compatible with healthy ecosystems. Natural ecosystems generally benefit from the variability of natural hydrological regimes. Other uses prefer less variability. Much of our engineering infrastructure is operated so as to reduce hydrological variability. Today water resource systems are increasingly required to provide rather than reduce hydrological (and accompanying sediment load) variability. Reservoir operators, for example, can modify their water release policies to increase this variability. Farmers and land-use developers must minimize rather than encourage land-disturbing activities. Floodplains may need to get wet occasionally. Rivers and streams may need to meander and fish species that require habitats along the full length of rivers to complete their life cycles must have access to those river reaches. Clearly these ecological objectives, added to all the other economic and environmental ones, can only compound the conflicts and issues with respect to land and water management and use.

So, how can we manage all this conflict and uncertainty? We know that water resources planning and management should be founded on sound science, efficient public programme administration and the broad participation of stakeholders. Yet obtaining each of these three conditions is a challenge. While the natural and social sciences can help us predict the economic, environmental and ecological impacts of alternative decisions, those predictions are never certain. In addition, these sciences offer no help in determining the best decision to make in the face of multiple conflicting goals held by multiple stakeholders – goals that have changed, and no doubt will continue to change. Water resources planning and management and decision-making is not as easy as 'we professionals can tell you what to do, all you need is the will to do it'. Very often it is not clear what should be done. Professionals administering the science, often from public agencies, non-governmental organizations, or even from universities, are merely among all the stakeholders having an interest in and contributing to the management of water.

Each governmental agency, consulting firm, environmental interest group and citizen typically has particular limitations, authorities, expertise and conflicts with other people, agencies and organizations, all tending to detract from achieving a fully integrated approach to water resources planning and management. But precisely because of this, the participation and contributions of all these stakeholders are needed. They must come together in a partnership if indeed an integrated approach to water resources planning and management is to be achieved and sustained. All views must be heard, considered and acted upon by all involved in the water resources planning and management process.

Water resources planning and management is not simply the application and implementation of science. It is creating a social environment that brings in all of us who should be involved, from the beginning, in a continuing planning process. This process is one of:

- educating ourselves about how our systems work and function
- identifying existing or potential options and opportunities for enhancement and resource development and use

- resolving the inevitable problems and conflicts that will result over who gets what and when, and who pays who for what and when and how much
- making and implementing decisions, and finally of
- monitoring the impacts of those decisions.

This process is repeated as surprises or new opportunities or new knowledge dictates.

Successful water resources planning and management requires the active participation of all community institutions involved in economic development and resource management. How can this begin at the local stakeholder level? How does anyone get others interested in preventing problems before those problems are apparent, and especially before 'unacceptable' solutions are offered to deal with them? And how do you deal with the inevitable group or groups of stakeholders who see it in their best interest not to participate in the planning process, but simply to criticize it from the outside? Who is in a position at the local level to provide the leadership and financial support needed? In some regions, non-governmental institutions have been instrumental in initiating and coordinating this process at local grass-root levels.

Water resources planning and management processes should identify a vision that guides development and operational activities in the affected region. Planning and management processes should:

- recognize and address the goals and expectations of the region's stakeholders
- identify and respond to the region's water-related problems
- function effectively within the region's legal/institutional frameworks
- accommodate both short and long-term issues
- generate a diverse menu of alternatives
- integrate the biotic and abiotic parts of the basin
- take into account the allocation of water for all needs, including those of natural systems
- be stakeholder driven
- take a global perspective
- be flexible and adaptable
- drive regulatory processes, not be driven by them
- be the basis for policy making
- foster coordination among planning partners and consistency among related plans
- be accommodating of multiple objectives
- be a synthesizer, recognize and deal with conflicts
- produce recommendations that can be implemented.

All too often integrated planning processes are hampered by the separation of planning, management and implementing authorities, turf-protection attitudes, shortsighted focusing of efforts, lack of objectivity on the part of planners, and inadequate funding. These deficiencies need addressing if integrated holistic planning and management is to be more than just something to write about.

Effective water resources planning and management is a challenge today, and will be an increasing challenge into the foreseeable future. This book introduces some of the tools that are being used to meet these challenges. We consider it only a step towards becoming an accomplished planner or manager.

## 7. References

ASCE (AMERICAN SOCIETY OF CIVIL ENGINEERS). 1998. *Sustainability criteria for water resource systems*. Reston, Va., ASCE.

GWP (GLOBAL WATER PARTNERSHIP). 2000. *Integrated water resources management*. Tac Background Papers No. 4. Stockholm, Sweden, GWP.

NRC (NATIONAL RESEARCH COUNCIL). 2001. *Compensating for wetland losses under the clean water act*. Committee on Mitigating Wetland Losses, Board on Environmental Studies and Toxicology, Water Science and Technology Board.

NATIONAL RESEARCH COUNCIL. 1999. Water for the future: the West Bank and Gaza Strip, Israel, and Jordan. Water Science and Technology Board and the Board on Environmental Studies and Toxicology, National Research Council, National Academy Press, Washington, DC.

WCD. 2000. *Dams and Developments – A new framework for decision-making: the report of the World Commission on Dams*. UK, Earthscan.

### Additional References (Further Reading)

ABU-ZEID, M.A. and BISWAS, A.K. (eds.). 1996. *River basin planning and management*. Oxford University Press, Calcutta.

BARROW, C.J. 1998. River basin development planning and management: a critical review. *World-Development*, Vol. 26, No. 1, pp. 171–86.

BISWAS, A.K. (ed.). 1997. *Water resources: environmental planning, management, and development*. New York, McGraw-Hill.

COOPER, A.B. and BOTTCHER, A.B. 1993. Basin-scale modelling as a tool for water-resource planning. *Journal of Water Resources Planning and Management (ASCE)*, Vol. 119, No. 3, pp. 306–23.

DIAMANTINI, C. and ZANON, B. 1996. River basin planning in Italy: resource and risk management. *European-Environment*, Vol. 6, No. 4, pp. 119–25.

ECKSTEIN, O. 1958. *Water resource development: the economics of project evaluation*. Cambridge, Mass., Harvard University Press.

GLOBAL WATER PARTNERSHIP (GWP). 2000. *Water as a social and economic good: how to put the principle into practice*. Stockholm, Sweden, GWP, Tac Background Papers No. 2.

GLOBAL WATER PARTNERSHIP (GWP). 2000. *Effective water governance*. Stockholm, Sweden, GWP, Tac Background Papers No. 7.

GOULTER, I.C. 1985. Equity issues in the implementation of river basin planning. In: J. Lundqvist (ed.), *Strategies for river basin management: environmental integration of land and water in a river basin*, pp. 287–92. Dordrecht, Holland, D. Reidel.

HOWE, C.W. 1996. Water resources planning in a federation of states: equity versus efficiency. *Natural Resources Journal*, Vol. 36, No. 1, pp. 29–36.

HUAICHENG, G. and BEANLANDS, G. 1994. A comparative study on Canadian and Chinese river basin planning. *Journal of Environmental Science China*, Vol. 6, No. 2, pp. 224–33.

KARAMOUZ, M.; ZINSSER, W.K. and SZIDAROVSZKY, F. 2003. *Water resources systems analysis*. Boca Raton, Fla., Lewis.

KRUTILLA, J.V. and ECKSTEIN, O. 1958. *Multiple purpose river development*. Baltimore, Md., Johns Hopkins Press.

KULSHRESHTHA, S. 1998. A global outlook for water resources to the year 2025. *Water Resources Management*, Vol. 12, No. 3, June, pp. 167–84.

LEE, D.J. and DINAR, A. 1996. An integrated model of river basin planning, development, and management. *Water International,* Vol. 21, No. 4, pp. 213–22. Also see 1995. *Review of integrated approaches to river basin planning, development and management*, Washington, D.C., World Bank, Agriculture and Natural Resources Department.

LINS, H.F.; WOLOCK, D.M. and MCCABE, G.J. 1997. Scale and modelling issues in water resources planning. *Climate Change*, Vol. 37, No. 1, pp. 63–88.

LOUCKS, D.P. (ed.). 1998. *Restoration of degraded rivers: challenges, issues and experiences*. Dordrecht, Holland, Kluwer Academic.

LOUCKS, D.P. and DA COSTA, J.R. (eds.). 1991. *Decision support systems: water resources planning and research*. Berlin, Springer-Verlag.

LOUCKS, D.P.; STEDINGER, J.R. and HAITH, D.A. 1981. *Water resources systems planning and analysis*. Englewood Cliffs, N.J., Prentice-Hall.

MAASS, A.; HUFSCHMIDT, M.M.; DORFMAN, R.; THOMAS, H.A. Jr.; MARGLIN, S.A. and FAIR, G.M. 1962. *Design of water resource systems*. Cambridge, Mass., Harvard University Press.

MAIDMENT, D.R. (ed.). 1993. *Handbook of hydrology*. New York, McGraw-Hill.

MAJOR, D.C. and LENTON, R.L. 1979. *Applied water resource systems planning*. Englewood Cliffs, N.J., Prentice-Hall.

MAYS, L.W. (ed.). 1996. *Water resources handbook*. New York, McGraw-Hill.

MCMILLAN, T. 1990. Water resource planning in Canada. *Journal of Soil and Water Conservation*, Vol. 45, No. 6, November/December, pp. 614–16.

MITCHELL, B. 1983. Comprehensive river basin planning in Canada: problems and opportunities. *Water International*, Vol. 8, No. 4, pp. 146–53.

O'RIORDAN, J. 1981. New strategies for water resource planning in British Columbia. *Canadian Water Resources Journal*, Vol. 6, No. 4, pp. 13–43.

RAZAVIAN, D.; BLEED, A.S.; SUPALLA, R.J. and GOLLEHON, N.R. 1990. Multistage screening process for river basin planning. *Journal of Water Resources Planning and Management (ASCE)*, Vol. 116, No. 3, May/June , pp. 323–34.

REITSMA, R.F. and CARRON, J.C. 1997. Object-oriented simulation and evaluation of river basin operations. *Journal of Geographic Information and Decision Analysis*, Vol. 1, No. 1, pp. 9–24.

REYNOLDS, P.J. Ecosystem approaches to river basin planning strategies for river basin management. In: J. Lundqvist (ed.), 1985. *Environmental integration of land and water in a river basin,* pp. 41–8. Dordrecht, Holland, D. Reidel.

SAHA, S.K. and BARROW, C.J. (eds.). 1981. *River basin planning: theory and practice*. Chichester, UK, Wiley Interscience.

SAVENIJE, H.H.G. and VAN DER ZAAG, P. (eds.). 1998. *The Management of Shared River Basins*. Neda, The Hague, Ministry of Foreign Affairs.

SCHRAMM, G. 1980. Integrated river basin planning in a holistic universe. *Natural Resources Journal*, Vol. 20, No. 4, October, pp. 787–806.

SMITH, S.C. and CASTLE E.N. (eds.). 1964. *Economics and public policy in water resources development*. Ames, Iowa, Iowa University Press.

SOMLYODY, L. Use of optimization models in river basin water quality planning. In: M.B. Beck and P. Lessard (eds.). 1997. *WATERMATEX '97: systems analysis and computing in water quality management. Towards a New Agenda*, pp. 73–87. London, International Water Association.

STOUT, G.E. 1998. Sustainable development requires the full cooperation of water users. *Water International*, Vol. 23, No. 1, March, pp. 3–7.

THANH, N.C. and BISWAS, A.K. (eds.). *Environmentally-sound water management*. Delhi, Oxford University Press.

THIESSEN, E.M.; LOUCKS, D.P. and STEDINGER, J.R. 1998. Computer-assisted negotiations of water resources conflicts. *Group Decision and Negotiation*, Vol. 7, No. 2, pp. 109–29.

VIESSMAN, W. 1996. Integrated water management. *Water Resources Update*, No. 106, Winter, pp. 2–12.

VIESSMAN, W. 1998. Water Policies for the Future. *Water Resources Update*, No. 111, Spring, pp. 4–7, 104–10.

# 2. Water Resource Systems Modelling: Its Role in Planning and Management

1. Introduction 39

2. Modelling of Water Resources Systems 41
    2.1. An Example Modelling Approach 41
    2.2. Characteristics of Problems to be Modelled 41

3. Challenges in Water Resources Systems Modelling 43
    3.1. Challenges of Planners and Managers 43
    3.2. Challenges of Modelling 44
    3.3. Challenges of Applying Models in Practice 45

4. Developments in Modelling 46
    4.1. Modelling Technology 46
    4.2. Decision Support Systems 47
        4.2.1. Shared-Vision Modelling 49
        4.2.2. Open Modelling Systems 51
        4.2.3. Example of a DSS for River Flood Management 51

5. Conclusions 54

6. References 55

# 2 Water Resource Systems Modelling: Its Role in Planning and Management

Planning, designing and managing water resources systems today inevitably involve impact prediction. Impact prediction involves modelling. While acknowledging the increasingly important role of modelling in water resources planning and management, we also acknowledge the inherent limitation of models as representations of any real system. Model structure, input data, objectives and other assumptions related to how the real system functions or will behave under alternative infrastructure designs and management policies or practices may be controversial or uncertain. Future events are always unknown and of course any assumptions about them may affect model outputs, that is, their predictions. As useful as they may or may not be, the results of any quantitative analysis are always only a part, but an important part, of the information that should be considered by those involved in the overall planning and management decision-making process.

## 1. Introduction

When design and management decisions are made about environmental and water resources systems, they are based on what the decision-makers believe, or perhaps hope, will take place as a result of their decisions. These predictions are either based on very qualitative information and beliefs in peoples' heads – or crystal balls (Figure 2.1) – or, at least in part, on quantitative information provided by mathematical or computer-based models (Figure 2.2). Today computer-based modelling is used to enhance mental models. These quantitative mathematical models are considered essential for carrying out environmental impact assessments. Mathematical simulation and optimization models packaged within interactive computer programs provide a common way for planners and managers to predict the behaviour of any proposed water resources system design or management policy before it is implemented.

Modelling provides a way, perhaps the principal way, of predicting the behaviour of proposed infrastructural designs or management policies. The past thirty years have witnessed major advances in our abilities to model the engineering, economic, ecological, hydrological and sometimes even the institutional or political impacts of large, complex, multipurpose water resources systems. Applications of models to real systems have improved our understanding, and hence have often contributed to improved system design, management and operation. They have also taught us how limited our modelling skills remain.

Water resources systems are far more complex than anything analysts have been, or perhaps ever will be, able to model and solve. The reason is not simply any computational limit on the number of model variables, constraints, subroutines or executable statements in those subroutines. Rather it is because we do not understand sufficiently the multiple interdependent physical, biochemical, ecological, social, legal and political

**Figure 2.1.** Using mental models for prediction.

**Figure 2.2.** Using computer models for prediction.

(human) processes that govern the behaviour of water resources systems. These processes are affected by uncertainties in things we can measure, such as water supply and water demands. They are also affected by the unpredictable actions of multiple individuals and institutions that are affected by what they get or do not get from the management and operation of such systems, as well as by other events having nothing directly to do with water.

The development and application of models – in other words, the art, science and practice of modelling, as will be discussed in the following chapters – should be preceded by the recognition of what can and cannot be achieved from the use of models. Models of real-world systems are always simplified representations. What features of the actual system are represented in a model, and what features are not, will depend in part on what the modeller thinks is important with respect to the issues being discussed or the questions being asked. How well this is done will depend on the skill of the modeller, the time and money available, and, perhaps most importantly, the modeller's understanding of the real system and decision-making process.

Developing models is an art. It requires knowledge of the system being modelled, the client's objectives, goals and information needs, and some analytical and programming skills. Models are always based on numerous assumptions or approximations, and some of these may be at issue. Applying these approximations of reality in ways that improve understanding and eventually lead to a good decision clearly requires not only modelling skills but also the ability to communicate effectively.

Models produce information. They do not produce decisions. Water resources planners and managers must accept the fact that decisions may not be influenced by their planning and management model results. To know, for example, that cloud seeding may, on average, reduce the strength of hurricanes over a large region does not mean that such cloud-seeding activities will or should be undertaken. Managers or operators may know that not everyone will benefit from what they would like to do, and those who lose will likely scream louder than those who gain.

In addition, decision-makers may feel safer in inaction than action (Shapiro, 1990; Simon, 1988). There is a strong feeling in many cultures and legal systems that failure to act (nonfeasance) is more acceptable than acts that fail (misfeasance or malfeasance). We all feel greater responsibility for what we do than for what we do not do. Yet our aversion to risk should not deter us from addressing sensitive issues in our models. Modelling efforts should be driven by the need for information and improved understanding. It is that improved understanding (not improved models per se) that may eventually lead to improved system design, management and/or operation. Models used to aid water resources planners and managers are not intended to be, and rarely are (if ever), adequate to replace their judgement. This we have learned, if nothing else, in over forty years of modelling experience.

This brief chapter serves as an overview of modelling and its applications. The emphasis is on application. This chapter is about modelling in practice more than in theory. It is based on the considerable experience and literature pertaining to how well, or how poorly, professional practitioners and researchers have done over the past four decades or more in applying various modelling approaches or tools to real problems with real clients (also see, for example, Austin, 1986; Gass, 1990; Kindler, 1987, 1988; Loucks et al., 1985; Reynolds, 1987 and Rogers and Fiering, 1986).

In attempting to understand how modelling can better support planners and managers, it may be useful to examine just what planners and managers of complex water resources systems do. What they do governs to some extent what they need to know. And what they need to know governs to a large extent what modellers or analysts should be trying to provide. In this book the terms analysts or modellers, planners, and managers can refer to the same person or group of individuals. The terms are used to distinguish the activities of individuals, not necessarily the individuals themselves.

First, a brief example is presented to demonstrate the value of modelling. Then we offer some general thoughts on the major challenges facing water resources systems planners and managers, the information they need to meet these challenges, and the role analysts have in helping to provide this information. Finally, we argue why we think the practice of modelling is in a state of transition, and how current research and development in modelling and computing technology are affecting that transition. New computer technology has had and will continue to have a significant impact in the development and use of models for water resources planning and management.

# 2. Modelling of Water Resources Systems

## 2.1. An Example Modelling Approach

Consider for example the sequence or chain of models required for the prediction of fish and shellfish survival as a function of nutrient loadings into an estuary. The condition of the fish and shellfish are important to the stakeholders. One way to maintain healthy stocks is to maintain sufficient levels of oxygen in the estuary. The way to do this is to control algae blooms. This in turn requires limiting the nutrient loadings to the estuary that can cause algae blooms and subsequent dissolved oxygen deficits. The modelling challenge is to link nutrient loading to fish and shellfish survival. In other words, can some quantitative relationship be defined relating the amount of nutrient loading to the amount of fish and shellfish survival?

The negative effects of excessive nutrients (e.g., nitrogen) in an estuary are shown in Figure 2.3. Nutrients stimulate the growth of algae. Algae die and accumulate on the bottom where bacteria consume them. Under calm wind conditions density stratification occurs. Oxygen is depleted in the bottom water. Fish and shellfish may die or become weakened and more vulnerable to disease.

A sequence of deterministic – or better, of probabilistic – models, each providing input data to the next model, can be defined (Chapter 12) to predict shellfish and fish abundance in the estuary based on upstream nutrient loadings. These models, for each link shown in Figure 2.4, can be a mix of judgemental, mechanistic and/or statistical ones. Statistical models could range from simple regressions to complex artificial neural networks (Chapter 6). Any type of model selected will have its advantages as well as its limitations, and its appropriateness may largely depend on the amount and precision of the data available for model calibration and verification.

The biological endpoints ‘shell-fish abundance’ and ‘number of fish-kills’ are meaningful indicators to stakeholders and can easily be related to designated water body use.

## 2.2. Characteristics of Problems to be Modelled

Problems motivating modelling and analyses exhibit a number of common characteristics. These are reviewed here because they provide insight into whether a modelling study of a particular problem may be worthwhile. If the planners’ objectives are very unclear, if few alternative courses of action exist, or if there is little scientific understanding of the issues involved, then mathematical modelling and sophisticated methodologies are likely to be of little use.

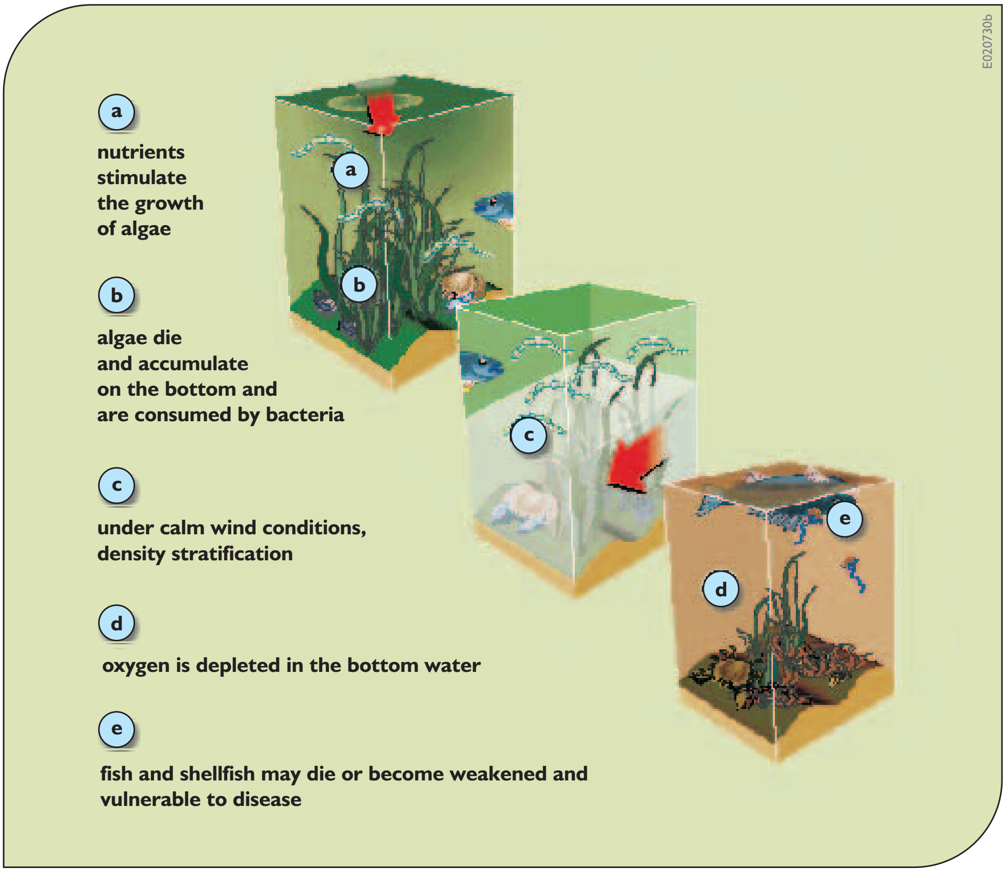



**Figure 2.3.** The impacts of excessive nutrients in an estuary (Borsuk et al., 2001).

Successful applications of modelling are often characterized by:

- *A systems focus or orientation*. In such situations attention needs to be devoted to the interdependencies and interactions of elements within the system as a whole, as well as to the elements themselves.
- *The use of interdisciplinary teams*. In many complex and non-traditional problems it is not at all clear from the start what disciplinary viewpoints will turn out to be most appropriate or acceptable. It is essential that participants in such work – coming from different established disciplines – become familiar with the techniques, vocabulary and concepts of the other disciplines involved. Participation in interdisciplinary modelling often requires a willingness to make mistakes at the fringes of one's technical competence and to accept less than the latest advances in one's own discipline.
- *The use of formal mathematics*. Most analysts prefer to use mathematical models to assist in system description and the identification and evaluation of efficient tradeoffs among conflicting objectives, and to provide an unambiguous record of the assumptions and data used in the analysis.

Not all water resources planning and management problems are suitable candidates for study using modelling methods. Modelling is most appropriate when:

- The planning and management objectives are reasonably well defined and organizations and individuals can be identified who can benefit from understanding the model results.
- There are many alternative decisions that may satisfy the stated objectives, and the best decision is not obvious.
- The water resources system and the objectives being analysed are describable by reasonably tractable mathematical representations.
- The information needed, such as the hydrological, economic, environmental and ecological impacts

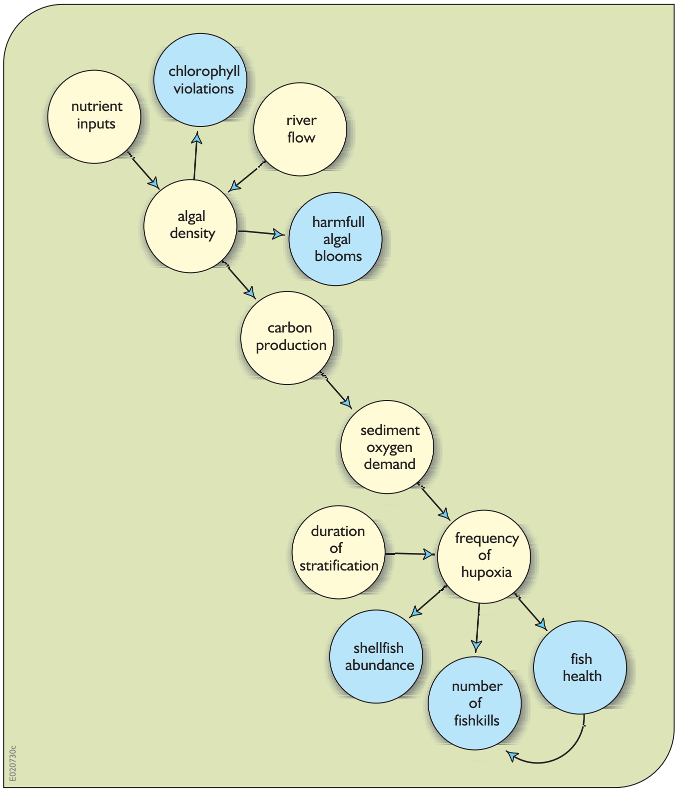



**Figure 2.4. Cause and effect diagram for estuary eutrophication due to excessive nutrient loadings (Borsuk et al., 2001).**

resulting from any decision, can be better estimated through the use of models.

- The parameters of these models are estimable from readily obtainable data.

# 3. Challenges in Water Resources Systems Modelling

## 3.1. Challenges of Planners and Managers

Planners and managers of water resources systems are the people responsible for solving particular water-related problems or meeting special water resources needs. When they fail, they hear about it. The public lets them know. What makes their job particularly challenging is that the public has differing needs and expectations. Furthermore, institutions where water resources planners and managers work (or hire consultants to work for them) are like most institutions these days: they must do what they can with limited financial and human resources. Their clients are all of us who use water, or at least all of us who are affected by the decisions they make.

The overall objective of these planners and managers and their institutions is to provide a service, such as a reliable and inexpensive supply of water, an assurance of water quality, the production of hydropower, protection from floods, the provision of commercial navigation and recreational opportunities, the preservation of wildlife and enhancement of ecosystems, or some combination of these or other purposes. Furthermore they are expected to do this at a cost no greater than what people are willing to pay. Meeting these goals (i.e., keeping everyone happy) is not always easy or even possible.

Simple technical measures or procedures are rarely able to ensure a successful solution to any particular set of water resources management problems. Furthermore, everyone who has had any exposure to water resources planning and management knows that one cannot design or operate a water resources system without making compromises. These compromises are over competing purposes (such as hydropower and flood control) or competing objectives (such as who benefits and who pays, and how much and where and when). After analysts, using their models of course, identify possible ways of achieving various goals and objectives and provide estimates of associated economic, environmental, ecological and social impacts, it is the planners and managers who have the more difficult job. They must work with and influence everyone who will be affected by whatever decision they make.

Planning and managing involves not only decision-making, but also developing among all interested and influential individuals an understanding and consensus that legitimizes the decisions and enhances their successful implementation. Planning and managing are processes that take place in a social or political environment. They involve leadership and communication among people and institutions, and the skills required are learned from experience of working with people, not with computers or models.

Moving an organization or institution into action to achieve specific goals involves a number of activities, including goal-setting, debating, coordinating, motivating, deciding, implementing and monitoring. Many of these activities must be done simultaneously and continuously, especially as conditions (goals and objectives, water supplies, water demands, financial budgets) change over time. These activities create a number of challenges that are relevant to modellers or analysts. They include how to:

- identify creative alternatives for solving problems
- find out what each interest group wants to know in order to reach an understanding of the issues and a consensus on what to do
- develop and use models and present their results so that everyone can reach a common or shared understanding and agreement that is consistent with their individual values
- make decisions and implement them, given differences in opinions, social values and objectives.

In addressing these needs or challenges, planners and managers must consider the relevant:

- legal rules and regulations
- history of previous decisions
- preferences of important actors and interest groups
- probable reactions of those affected by any decision
- relative importance of various issues being addressed
- applicable science, engineering and economics – the technical aspects of their work.

We mention these technical aspects lastly not to suggest that they are the least important factor to be considered. We do so to emphasize that they are only one set among many factors and probably, in the eyes of planners and managers, not the most decisive or influential (Ahearne, 1988; Carey, 1988; Pool, 1990 and Walker, 1987).

So, does the scientific, technical, systematic approach to modelling for planning and management really matter? We believe it can if it addresses the issues of concern to the modellers' clients: the planners and the managers. Analysts need to be prepared to interact with the political or social structure of the institutions they are attempting to assist, as well as with the public and the press. Analysts should also be prepared to have their work ignored. Even if they are presenting 'facts' based on the current state of the sciences, these sciences may not be considered relevant. Fortunately for scientists and engineers, this is not always the case. The challenge of modellers or analysts interested in having an impact on the practice of water resources systems planning and management is to become a part of the largely political planning and management process and to contribute towards its improvement.

## 3.2. Challenges of Modelling

To engage in a successful water resources systems study, the modeller must possess not only the requisite mathematical and systems methodology skills, but also an understanding of the environmental engineering, economic, political, cultural and social aspects of water resources planning problems. Consider, for example, the study of a large land-development plan. The planner should be able to predict how the proposed development will affect the quantity and quality of the surface and subsurface runoff, and the impact this will have on the quantity and quality of surface and ground waters and their ecosystems. These impacts, in turn, might affect the planned development itself, or other land uses downstream. To do this the analysts must have an understanding of the biological, chemical, physical and even social processes that are involved in water resources management.

A reasonable knowledge of economic theory, law, regional planning and political science can be just as important as an understanding of hydraulic, hydrogeologic, hydrological, ecologic and environmental engineering disciplines. It is obvious that the results of most water resources management decisions have a direct impact on people and their relationships. Hence inputs from those having a knowledge of those other disciplines are also needed during the comprehensive planning of water resources systems, especially during the development and evaluation of the results of various planning models.

Some of the early water resources systems studies were undertaken with a naïve view of the appropriate role and impact of models and modellers in the policy-making process. The policy-maker could foresee the need to make a decision. He or she would ask the systems group to study the problem. They would then model it, identify feasible solutions and their consequences, and recommend one or at most a few alternative solutions. The policy-maker, after waiting patiently for these recommendations, would then make a yes or no decision. However, experience to date suggests the following:

- A final solution to a water resources planning problem rarely exists: plans and projects are dynamic. They evolve over time as facilities are added and modified to adapt to changes in management objectives and in the demands placed on the facilities.

- For every major decision there are many minor decisions, made by different agencies or management organizations responsible for different aspects of a project.
- The time normally available to study particular water resources problems is shorter than the time needed; if there is sufficient time, the objectives of the original study will probably have shifted significantly by the time the study is completed.

This experience emphasizes some of the limitations and difficulties that any water resources systems study may encounter, but more importantly, it underscores the need for constant communication among the analysts, system planners, managers and operators, and policy-makers. The success or failure of many past water resource studies is due largely to the efforts expended or not expended in ensuring adequate, timely and meaningful communication – communication among systems analysts, planners, those responsible for system operation and design, and public officials responsible for major decisions and setting general policies. Decision-makers who need the information that can be derived from various models and analyses, need it at particular times and in a form useful and meaningful to them. Once their window of opportunity for decision-making has passed, such information, no matter how well presented, is often useless.

At the beginning of any study, objectives are usually poorly defined. As more is learned about what can be achieved, stakeholders are better able to identify what they want to do. Close communication among analysts and all interested stakeholders and decision-makers throughout the modelling process is essential if systems studies are to make their greatest contribution to the planning process. Objectives as stated at the beginning of a study are rarely the objectives as understood at its end.

Furthermore, those who will use models, and present the information derived from models to those responsible for making decisions, must be intimately involved with model development, solution and analysis. It is only then can they appreciate the assumptions upon which any particular model is based, and hence adequately evaluate the reliability of the results. A water resources systems study that involves only outside consultants, and has minimal communication between consultants and planners within a responsible management agency or involved stakeholders, is unlikely to have a significant impact on the planning process. Models that are useful are constantly being modified and applied by those involved in plan preparation, evaluation and implementation.

The interaction described above is illustrated in Figure 1.20 of the previous chapter. Models are developed and applied during the second and third phase of this analytical framework. A continuous communication with the decision-makers and stakeholder representatives should ensure the models and results will indeed serve their purpose.

## 3.3. Challenges of Applying Models in Practice

As already mentioned, the clients of modellers or analysts are typically planners and managers who have problems to solve and who could benefit from a better understanding of what options they have and what impacts may result. They want advice on what to do and why, what will happen as a result of what they do, and who will care and how much. The aim of analysts is to provide planners and managers with meaningful (understandable), useful, accurate and timely information. This information serves to help them better understand their system, its problems, and alternative ways to address them. The purpose of water resources systems planning and management modelling, stated once again, is to provide useful and timely information to those involved in managing such systems.

Modelling is a process or procedure intended to focus and force clearer thinking and to promote more informed decision-making. The approach involves problem recognition, system definition and bounding, identification of various goals or objectives, identification and evaluation of various alternatives, and very importantly, effective communication of this information to those who need to know.

The focus of most books and articles on water resource systems modelling is on modelling methods. This book is no different. But what all of us should also be interested in, and discuss more than we do, is the use of these tools in the processes of planning and management. If we did, we could learn much from each other about what tools are needed and how they can be better applied in practice. We could extend the thoughts of those who, in a more general way, addressed these issues over two decades ago (Majoni and Quade, 1980; Miser, 1980; Stokey and Zeckhauser, 1977 and Tomlison, 1980).

There is always a gap between what researchers in water resources systems modelling produce and publish, and what the practitioner finds useful and uses. Those involved in research are naturally interested in developing new and improved tools and methods for studying, identifying and evaluating alternative water resources system designs and management and operation policies. If there were no gap between what is being developed or advocated by researchers and that which is actually used by practitioners, either the research community would be very ineffective in developing new technology or the practitioners would be incredibly skilled in reading, assimilating, evaluating and adapting this research to meet their needs. Evaluation, testing and inevitable modifications take time. Not all published research is ready or suited for implementation. Some research results are useful, some are not. It is a work in progress.

How can modellers help reduce the time it takes for new ideas and approaches to be used in practice? Clearly, practitioners are not likely to accept a new modelling approach, or even modelling itself, unless it is obvious that it will improve the performance of their work as well as help them address problems they are trying to solve. Will some new model or computer program make it easier for practitioners to carry out their responsibilities? If it will, there is a good chance that the model or computer program might be successfully used, eventually. Successful use of the information derived from models or programs is, after all, the ultimate test of the value of those tools. Peer review and publication is only one, and perhaps not even a necessary, step towards that ultimate test or measure of value of a particular model or modelling approach.

# 4. Developments in Modelling

## 4.1. Modelling Technology

The increasing developments in computer technology – from microcomputers and workstations to supercomputers – have motivated the concurrent development of an impressive set of new models and computer software. This software is aimed at facilitating model use and, more importantly, interaction and communication between the analysts or modellers and their clients. It includes:

- interactive approaches to model operation that put users more in control of their computers, models, and data
- computer graphics that facilitate data input, editing, display and comprehension
- geographic information systems that provide improved spatial analysis and display capabilities
- expert systems that can help the user understand better how complex decision problems might be solved, and at the same time explain to the users why one particular decision may be better than another
- electronic mail and the Internet, which let analysts, planners and managers communicate and share data and information with others worldwide, and to run models that are located and maintained at distant sites
- multimedia systems that permit the use of sound and video animation in analyses, all aimed at improving communication and understanding.

These and other software developments are giving planners and managers improved opportunities for increasing their understanding of their water resources systems. Such developments in technology should continue to aid all of us in converting model output data to information; in other words, it should provide us with a clearer knowledge and understanding of the alternatives, issues and impacts associated with potential solutions to water resources systems problems. But once again, this improved information and understanding will only be a part of everything planners and managers must consider.

Will all the potential benefits of new technology actually occur? Will analysts be able to develop and apply these continual improvements in new technology wisely? Will we avoid another case of oversell or unfulfilled promises? Will we avoid the temptation of generating fancy animated, full-colour computer displays just because we are easily able to produce them, rather than working on the methods that will add to improved understanding of how to solve problems more effectively? Will we provide the safeguards needed to ensure the correct use and interpretation of the information derived from increasingly user-friendly computer programs? Will we keep a problem-solving focus, and continue to work towards increasing our understanding of how to improve the development and management of our water resources, whether or not our planning models are incorporated into some sort of interactive computer-aided support system? We can, but it will take discipline.

As modellers or researchers, we must discipline ourselves to work more closely with our clients: the planners, managers and other specialists who are responsible for the development and operation of our water resources systems. We must study their systems and their problems, and we must identify their information needs. We must develop better tools that they themselves and other interested stakeholders can use to model their water resource systems and obtain an improved understanding – a shared vision – of how their system functions and of their available management options and associated impacts or consequences. We must be willing to be multidisciplinary and capable of including all relevant data in our analyses. We must appreciate and see the perspectives of the agronomists, ecologists, economists, engineers, hydrologists, lawyers or political and regional scientists as appropriate. Viewing a water resources system from a single-discipline perspective is rarely sufficient for today's water resource systems planning.

Even if we have successfully incorporated all relevant disciplines and data in our analyses, we should have a healthy scepticism about our resulting information. We must admit that this information, especially concerning what might happen in the future, is uncertain. If we are looking into the future (whether using crystal balls as shown in Figure 2.1 or models as in Figure 2.2), we must admit that many of our assumptions, such as parameter values, cannot even be calibrated, let alone verified. Our conclusions or estimates can be very sensitive to those assumptions. One of our major challenges is to communicate this uncertainty in understandable ways to those who ask for our predictions.

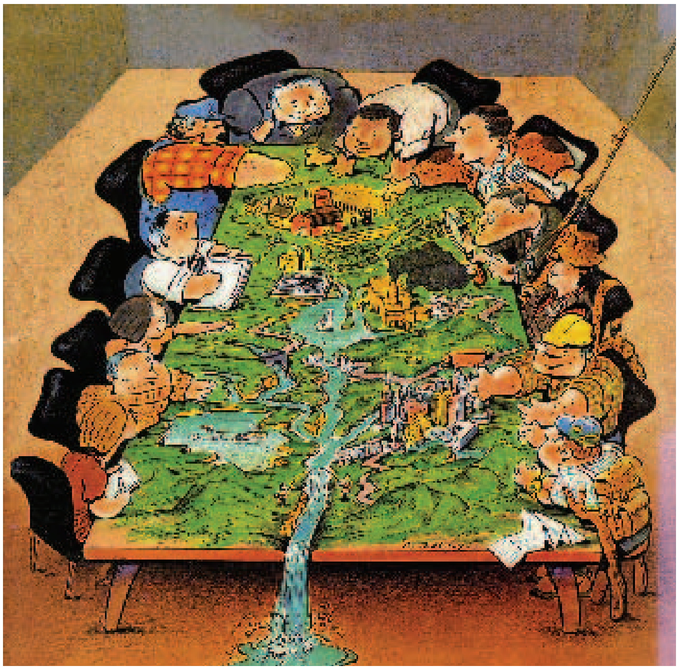


**Figure 2.5.** Stakeholders involved in river basin planning and management, each having different goals and information needs (*Engineering News Record*, 20 September 1993, with permission).

## 4.2. Decision Support Systems

Water resources planners and managers today must consider the interests and goals of numerous stakeholders. The planning, managing and decision-making processes involve negotiation and compromise among these numerous stakeholders, like those shown in Figure 2.5, who typically have different interests, objectives and opinions about how their water resources system should be managed. How do we model to meet the information needs of all these different stakeholders? How can we get them to believe in and accept these models and their results? How do we help them reach a common – shared – vision? How can we help create a shared vision among all stakeholders of at least how their system works and functions, if not how they would like it to?

Today we know how to build some rather impressive models of environmental systems. We know how to incorporate within our models the essential biology, chemistry and physics that govern how the environmental system works. We have also learned a little about how to include the relevant economics, ecology and engineering into these models. Why do we do this? We do all this modelling simply to be able to estimate, or identify, and compare and evaluate the multiple impacts resulting from different design and management decisions we might make. Such information, we assume, should be of value to those responsible for choosing the 'best' decision.

If our goal is to help prevent, or contribute to the solution of, water resources problems, then simply having information from the world's best models and technology, as judged by our peers, is not a guarantee of success. To be useful in the political decision-making process, the

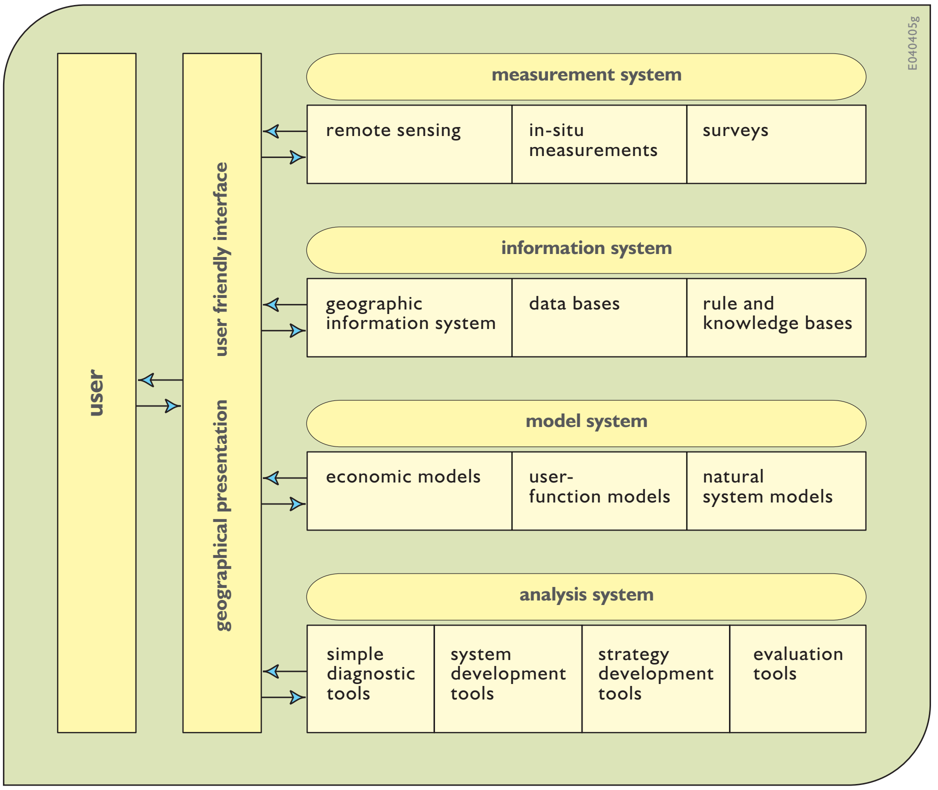



**Figure 2.6.** Common components of many decision support systems.

information we generate with all our models and computer technology must be understandable, credible and timely. It must be just what is needed when it is needed. It must be not too little nor too much.

The optimal format and level of detail and precision of any information generated from models should depend on the needs and backgrounds of each individual involved in the decision-making process. The value of such information, even if the format and content are optimal, will also depend on when it is available. Information on an issue is only of value if it is available during the time when the issue is being considered – that is, when there is an interest in that issue and a decision concerning what to do about it has not yet been made. That is the window of opportunity when information can have an impact. Information is of no value after the decision is made, unless of course it results in opening up another window of opportunity.

If there is truth in the expression 'decision-makers don't know what they want until they know what they can get', how do modellers know what decision-makers will need before even they do? How will modellers know what is the right amount of information, especially if they are to have that information available, and in the proper form, before or at the time, not after, it is needed? Obviously modellers cannot know this. However, over the last two decades or so this challenge has been addressed by developing and implementing decision support systems (DSSs) (Fedra, 1992; Georgakakos and Martin, 1996; Loucks and da Costa, 1991). These interactive modelling and display technologies can, within limits, adapt to the level of information needed and can give decision-makers some control over data input, model operation and data output. But will each decision-maker, each stakeholder, trust the model output? How can they develop any confidence in the models contained in a DSS? How can they modify those models within a DSS to address issues the DSS developer may not have considered? An answer to these questions has been the idea of involving the decision-makers themselves not only in interactive model use, but in interactive model building as well.

Figure 2.6 gives a general view of the components of many decision support systems. The essential feature is the interactive interface that permits easy and meaningful data entry and display, and control of model (or computer) operations.

**Various Phases of Decision Support Systems**

<table>
<tr><th></th><th>data provided by</th><th>data analysed by</th><th>options generated by</th><th>decision selection by</th><th>decision implemented by</th><th>approach to decision-making</th></tr>
<tr><td>1</td><td colspan="5">decision-maker</td><td>completely unsupported</td></tr>
<tr><td>2</td><td>GIS / DB</td><td colspan="4">decision-maker</td><td>information supported</td></tr>
<tr><td>3</td><td>GIS / DB</td><td>MODEL</td><td colspan="3">decision-maker</td><td>systematic analysis</td></tr>
<tr><td>4</td><td>GIS / DB</td><td colspan="2">MODEL</td><td colspan="2">decision-maker</td><td>sys. analysis alternatives</td></tr>
<tr><td>5</td><td>GIS / DB</td><td colspan="3">MODEL</td><td>decision-maker</td><td>system with over-ride</td></tr>
<tr><td>6</td><td>GIS / DB</td><td colspan="4">MODEL</td><td>automated</td></tr>
</table>

E040325c

**Figure 2.7.** Various types of computer-aided decision support systems (based on O'Callaghan, 1996).

Depending on the particular issue at hand, and more importantly the particular individuals and institutions involved, a decision support system in the broadest sense can range from minimal if any computer model use – where the decision-makers provide all the data and analyses, make the decision, and they or their institutions implement those decisions – to decision support systems that are fully automated and where no human involvement is present. The latter are rare, but they do exist. The automatic closing of the flood gates in Rotterdam harbour is an example of this. These extremes, and various levels of DSS in between are outlined in Figure 2.7.

### 4.2.1. Shared-Vision Modelling

Involving stakeholders in model building gives them a feeling of ownership. They will have a much better understanding of just what their model can do and what it cannot. If they are involved in model-building, they will know the assumptions built into their model. Being involved in a joint modelling exercise is a way to better understand the impacts of various assumptions. While there may be no agreement on the best of various assumptions to make, stakeholders can learn which of those assumptions matter and which do not. In addition, just the process of model development by numerous stakeholders will create discussions that can lead toward a better understanding of everyone's interests and concerns. Through such model-building exercises, it is just possible those involved will reach not only a better understanding of everyone's concerns, but also a common or 'shared' vision of at least how their water resources system works (as represented by their model, of course). Experience in stakeholder involvement in model-building suggests such model-building exercises can also help multiple stakeholders reach a consensus on how their real system should be developed and managed.

In the United States, one of the major advocates of shared vision modelling is the Institute for Water Resources of the US Army Corps of Engineers. They have applied their interactive general-purpose model-building platform in a number of exercises where conflicts existed over the design and operation of water systems (Hamlet, et al., 1996a, 1996b, 1996c; Palmer, Keys and Fisher, 1993; Werick, Whipple and Lund, 1996). Each of these model-building 'shared-vision' exercises included numerous stakeholders together with experts in the use of the software. Bill Werick of the Corps writes:

> Because experts and stakeholders can build these models together, including elements that interest each group, they gain a consensus view of how the water system

works as a whole, and how it affects stakeholders and the environment. Without adding new bureaucracies or reassigning decision-making authority, the shared vision model and the act of developing it create a connectedness among problem solvers that resembles the natural integration of the conditions they study.

Now the question is how to get all the stakeholders, many of whom may not really want to work together, involved in a model-building exercise. This is our challenge!

One step in that direction is the development of improved technologies that will facilitate model development and use by stakeholders with various backgrounds and interests. We need better tools for building DSSs, not just better DSSs themselves. We need to develop better modelling environments that people can use to make their own models. Researchers need to be building the model building blocks, as opposed to the models themselves, and to focus on improving those building blocks that can be used by others to build their own models. Clearly if stakeholders are going to be involved in model-building exercises, it will have to be an activity that is enjoyable and require minimal training and programming skills.

Traditional modelling experiences seem to suggest that there are five steps in the modelling process. The first is to identify the information the model is to provide. This includes criteria or measures of system performance that are of interest to stakeholders. These criteria or measures are defined as functions of the behaviour or state of the system being modelled. Next, this behaviour needs to be modelled so the state of the system associated with any 'external' inputs can be predicted. This requires modelling the physical, chemical, biological, economic, ecological and social processes that take place, as applicable, in the represented system. Thirdly, these two parts are put together, along with a means of entering the 'external' inputs and obtaining in meaningful ways the outputs. Next, the model must be calibrated and verified or validated, to the extent it can. Only now can the model be used to produce the information desired.

This traditional modelling process is clearly not going to work for those who are not especially trained or experienced (or even interested) in these modelling activities. They need a model-building environment where they can easily create models that:

- they understand
- are compatible with available data
- work and provide the level and amount of information needed
- are easily calibrated and verified when possible
- give them the interactive control over data input, editing, model operation and output display that they can understand and that they need in order to make informed decisions.

The challenge in creating such model-building environments is to make them sufficiently useful and attractive that multiple stakeholders will want to use them. They will have to be understandable. They will have to be relatively easy and transparent, and even fun, to build. They must be capable of simulating and producing different levels of detail with regard to natural, engineering, economic and ecological processes that take place at different spatial and temporal scales. And they must require no programming and debugging by the users. Just how can this be done?

One approach is to develop interactive modelling 'shells' specifically suited to modelling environmental problems. Modelling shells are data-driven programs that become models once sufficient data have been entered into them.

There are a number of such generic modelling shells for simulating water resources systems. AQUATOOL (Andreu et al., 1991), RIBASIM (Delft Hydraulics, 2004), MIKE-BASIN (Danish Hydraulic Institute, 1997) and WEAP (Raskin et al., 2001) (Shown in Figure 2.8) are representative of interactive river–aquifer simulation shells that require the system to be represented by, and drawn in as, a network of nodes and links. Each node and link requires data, and these data depend on what that node or link represents, as well as what the user wants to get from the output. If what is of interest is the time series of quantities of water flowing, or stored, within the system as a result of reservoir operation and/or water allocation policies, then water quality data need not be entered, even though there is the capacity to model water quality. If water quality outputs are desired, then the user can choose the desired various water quality constituents. Obviously, the more different types of information desired or the greater spatial or temporal resolution desired in the model output, the more input data required.

Interactive shells provide an interactive and adaptive way to define models and their input data. Once a model

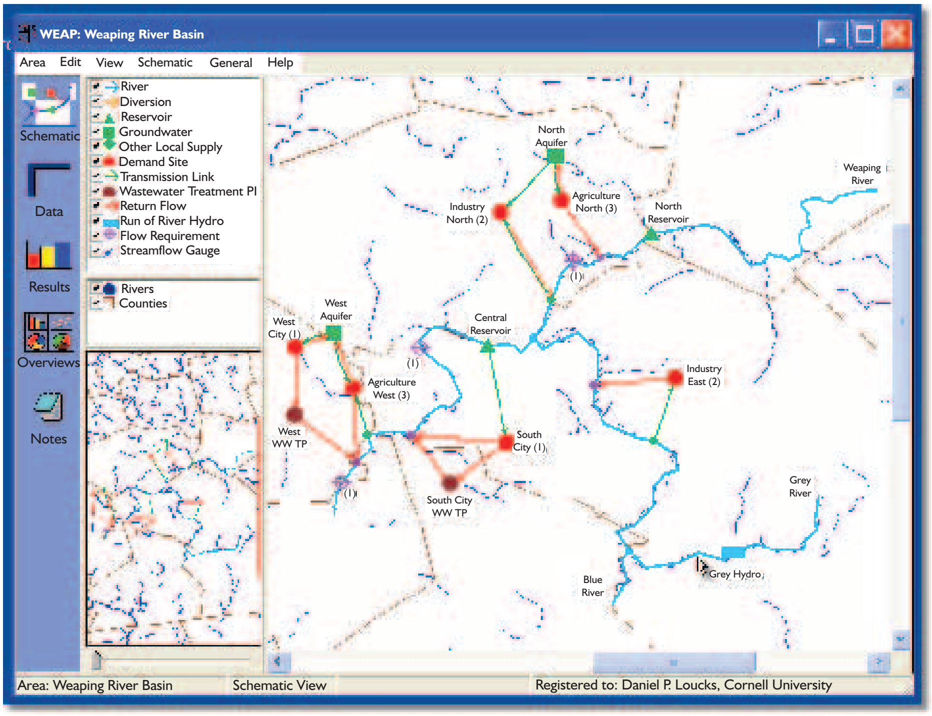



**Figure 2.8.** The main interface of the WEAP program, which is typical of a variety of generic river basin models that are able to simulate any river system drawn into the computer and displayed on the computer terminal, as shown.

is defined, the shell provides the interface for input data entry and editing, model operation and output data display.

To effectively use such shells, some training is useful in the use of the shell and what it can and cannot do. The developers of such shells have removed the need to worry about database management, solving systems of equations, developing an interactive interface, preserving mass balances and continuity of flow, and the like. Any assumptions built into the shell should be readily transparent and acceptable to all before it is used in any shared-vision exercises.

### 4.2.2. Open Modelling Systems

The next step in shared-vision modelling will be to create a modelling environment that will enable all stakeholders to include their own models in the overall system description. Stakeholders tend to believe their own models more than those provided by governmental agencies or research institutes. Their own models include the data they trust, and are based on their own assumptions and views on how the system works. For example, in transboundary water resources issues, different countries may want to include their own hydrodynamic models for the river reaches in their country.

Various developments on open modelling systems are taking place in Europe and the United States, although most of them are still in a research phase. The implementation of the Water Framework Directive in Europe has stimulated the development of OpenMI (European Open Modelling Interface and Environment). OpenMI will simplify the linking of water-related models that will be used in the strategic planning required by the Water Framework Directive (Gijsbers et al., 2002). An initiative in the United States aims to establish a similar framework for Environmental Models (Whelan and Nicholson, 2002).

### 4.2.3. Example of a DSS for River Flood Management

In the Netherlands the flood management policy on the Rhine Branches is aimed at reducing flood stages, since further raising of the dyke system is judged as unsustainable in the long term. Possible measures to reduce flood stages include the removal of hydraulic obstacles, lowering of groins, widening of the low-flow channel, lowering of floodplains, setting back of dykes, construction of side

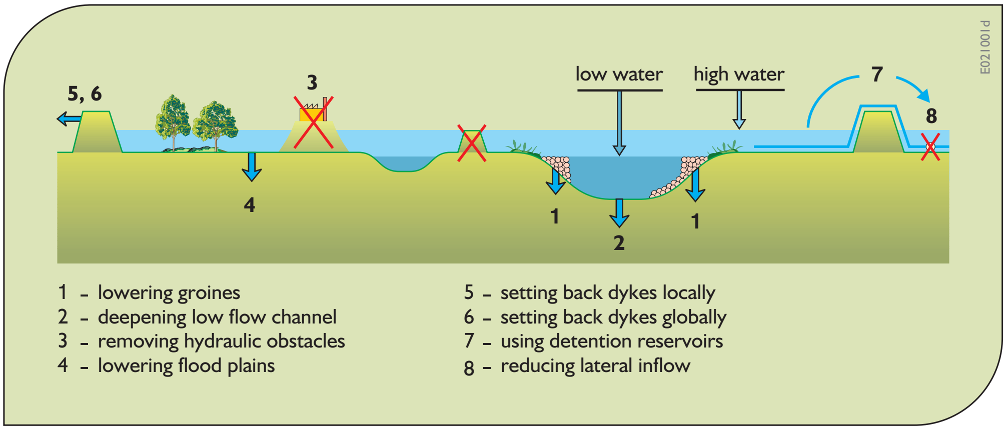



**Figure 2.9.** River improvement measures (see Appendix D).

channels, detention basins and other measures (as described in more detail in Appendix D). These options are illustrated in Figure 2.9.

Determining which set of river improvement measures to implement involves a complex process of public decision-making that includes many stakeholders. Exploratory investigations along the river have identified over 600 possible improvement measures. Which of these alternatives should be chosen? A decision needs to be made, and it needs to be acceptable to at least the majority of stakeholders. As this is being written this decision-making process is taking place. It is benefiting from the use of online decision support that provides information on the flood levels resulting from combinations of measures along the river. This relatively simple and user-friendly decision support system is called a Planning Kit. (This kit is available from Delft Hydraulics.)

The preliminary design phase of this scheme, called Room for the Rhine Branches, consists to a large extent of bottom-up public decision-making processes. Starting from the notion that multiple usage of space located between the river dykes (i.e., the area between the main embankments) should be possible. On the basis of a number of exploratory studies that identify possible measures and their respective effects, stakeholders and local authorities are to identify their preferred plans. These are to be judged on a number of criteria, such as the flood conveyance capacity of the river, its navigability, and its impact on the landscape and ecological infrastructure. The envisaged result of this procedure is an outline of a coherent scheme of river improvement.

The Planning Kit is developed for online decision support and to facilitate a public discussion – as well as one among professionals – in the planning and preliminary design phases.

Because of the large number of options and stakeholders, the selection process is complicated. However, reaching a technical optimum is not the objective. All of the river's functions, including its impact on the basin's ecology and cultural heritage, have to be respected. Public acceptance is an essential requirement. In the meantime, a number of overall criteria have to be satisfied. Without further support from a variety of models incorporated within the Planning Kit, this decision-making process would be much less directed or focused, and hence much less effective.

Numerical models of the Rhine exist. They vary in scale level (from basin-wide down to local scale) and in sophistication (1-D cross-sectionally averaged, 2-D depth-averaged, 2-D or 3-D eddy-resolving, etc.). In the studies in the framework of Room for the Rhine Branches, a 1-D model of the Lower Rhine is used for the large-scale phenomena and morphological computations and a 2-D depth-averaged model for more detailed local computations. Flood-level computations are made with a 2-D depth-averaged model of the entire Rhine Branches.

These models are being intensively used in the exploration and design phases of the river improvement works. They provide help in setting the target design water levels in order to avoid dyke raising, in checking the safety of the flood defences, and in assessing the hydraulic and morphological effects of proposed measures.

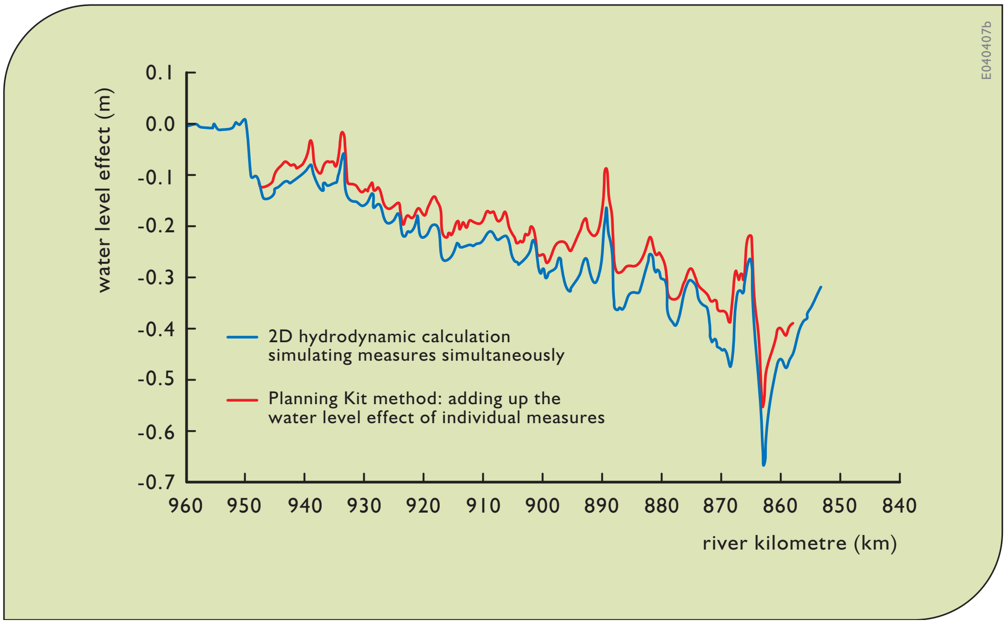



**Figure 2.10.** Water-level effect of forty measures along the Rhine Branch Waal under design flood conditions as calculated with a 2-D model simulating all measures (blue line) and as a result of the Planning Kit method in which results for individual measures are simply summed together.

The flood level computations with the 2-D model for the Rhine Branches are too time-consuming to be performed online during a meeting or a brainstorming session. Therefore, model runs for each of the suggested measures have been performed beforehand, and the water-level effects of individual measures have been stored in a database. This database is the core of the Planning Kit.

The basic assumption underlying the Planning Kit approach is that, as a first-order approximation, the water-level effects of each individual measure can be separated from those of every other measure, and the effects of combinations of measures can be obtained by superposition of their individual effects. At first sight, this may seem a disputable approach, since the hydrodynamic equations are essentially non-linear. Indeed, the total water-level effect of two combined measures may easily be 50% higher or lower than the sum of the two individual effects. But for large sets of measures (say, more than twenty-five over a 100-km river stretch) this approach has proven to be quite acceptable. As exemplified by Figure 2.10, a fully non-linear 2-D hydrodynamic model computation for a combination of forty measures along the Rhine Branch Waal results in water levels which are at most 10 cm lower than the results of the Planning Kit, in which water-level effects of individual measures are simply totalled.

The database approach allows for an online presentation of effects of measures. With the super-position principle, the design water-level effects of any combination of measures can be composed from those of each individual measure.

For each of the proposed measures, the database contains a situation sketch or an aerial photograph, ground photographs, the effects on the longitudinal water surface profile under the design flood conditions, the area covered, a cost estimate, the length of dyke to be rebuilt, ecological effects, amounts of material to be excavated for various soil types and so on. Thus, when analysing and synthesizing a set of measures, the aspects relevant to the decision-making can be immediately provided.

The central element in the presentation of this information (Figure 2.11) is a diagram showing the water-surface profile as far as it exceeds the desired profile according to the Room for the River principle (no dyke reinforcement). By clicking on a measure, it is activated and the water surface profile is adjusted accordingly. Thus, one can see immediately how much more is needed in order to reach the desired situation. By opening windows connected to the name of the measure, one can display photographs and any other available information.

The Planning Kit can easily be installed and run on a PC or a laptop computer; hence, it can conveniently be used online during meetings and hearings.

**Figure 2.11.** Main screen of the Planning Kit. The small geometric shapes are alternatives at various locations on the selected and displayed river reach. The plot shows the water levels associated with selected alternatives.

## 5. Conclusions

In our opinion, the most important aspect of model use today is communication. Unless water resources planners and managers can articulate well their needs for information, it will be difficult for modellers to generate such information. If the modellers cannot communicate effectively their modelling assumptions and results, or how others can use their tools to obtain their own results, then little understanding will be gained from such models. Both users and producers of modelling analyses must work together to improve communication. This takes time, patience and the willingness to understand what each has to say has well as the real meaning behind what is said.

To expect everyone to communicate effectively and to understand one another fully may be asking too much. There is a story written in the Bible (Genesis; Chapter 11, Verses 1–9) that tells us of a time when everyone on the earth was together and spoke one language. It seems these people decided to build a tower 'whose top may reach into the heaven'. Apparently this activity came to the attention of the Lord, who for some reason did not like this tower-building idea. So, according to the Bible, the Lord came down to earth and 'confounded the peoples language so they could not understand one another'. They could no longer work together to build their tower.

Is it any wonder we have to work so hard to communicate more effectively with one another, even in our single, but multidisciplinary, field of water resources planning and management? Let all of us modellers or analysts, planners and managers work together to build a new tower of understanding. To do this we need to control our jargon and take the time to listen, communicate and learn from each other and from all of our experiences.

Those who are involved in the development of water resources systems modelling methodology know that the use of these models cannot guarantee development of optimal plans for water resources development and management. Given the competing and changing objectives and priorities of different interest groups, the concept of an 'optimal plan' is not very realistic. What modellers can do, however, is to help define and evaluate, in a rather detailed manner, numerous alternatives that represent various possible compromises among conflicting groups, values and management objectives. A rigorous and objective analysis should help to identify the possible tradeoffs among quantifiable objectives so that further debate and analysis can be more informed. The art of modelling is to identify those issues and concerns that are important and significant and to structure the analysis to shed light on these issues.

Although water resources planning and management processes are not restricted to mathematical modelling, modelling is an important part of those processes. Models can represent in a fairly structured and ordered manner the important interdependencies and interactions among the various control structures and users of a water resources system. Models permit an evaluation of the consequences of alternative engineering structures, of various operating and allocating policies, and of different assumptions regarding future supplies, demands, technology, costs, and social and legal requirements. Although models cannot define the best objectives or set of assumptions, they can help identify the decisions that best meet any particular objective and assumptions.

We should not expect, therefore, to have the precise results of any quantitative systems study accepted and implemented. A measure of the success of any systems study resides in the answer to the following questions: Did the study have a beneficial impact in the planning and decision-making process? Did the results of such studies lead to a more informed debate over the proper choice of alternatives? Did it introduce competitive alternatives that otherwise would not have been considered?

There seems to be no end of challenging water resources systems planning problems facing water resources planners and managers. How one models any specific water resource problem depends on: first, the objectives of the analysis; second, the data required to evaluate the projects; third, the time, data, money and computational facilities available for the analysis; and fourth, the modeller's knowledge and skill. Model development is an art; it requires judgement both in abstracting from the real world the components that are important to the decision to be made and that can be illuminated by quantitative methods, and also in expressing those components and their inter-relationships mathematically in the form of a model. This art is to be introduced in the next chapter (Chapter 3).

## 6. References


AHEARNE, J.F. 1988. Addressing public concerns in science. *Physics Today*, Vol. 41, No. 9, pp. 36–42.

ANDREU, J.; CAPILLA, J. and SANCHIS, E. 1991. AQUATOOL. A computer assisted support system for water resources research management including conjunctive use. In: D.P. Loucks and J.R. da Costa (eds.), *Decision support systems*, Berlin, Springer-Verlag, pp. 333–55.

AUSTIN, T.A. 1986. Utilization of models in water resources. *Water Resources Bulletin*, Vol. 22, No. 1, pp. 49–56.

BORSUK, M.E.; STOW, C.A.; HIGDON, D. and RECKHOW, K.H. 2001. A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones. *Ecological Modelling*, No. 143, pp. 165–81.

CAREY, W.D. 1988. Scientists and sandboxes: regions of the mind. *American Scientist*, No. 76, pp. 143–5.

DANISH HYDRAULIC INSTITUTE (DHI). 1997. *MIKE-BASIN: operating manual and description*. Denmark, Hørsholm.

DELFT HYDRAULICS. 2004. *RIBASIM River basin simulation program operating manual and description*. Netherlands, Delft.

FEDRA, K. 1992. Advanced computer applications. *Options*, December.

GASS, S.I. 1990. Model world: 1-lave model, will travel. *Interfaces*, Vol. 20, No. 2, pp. 67–71.

GEORGAKAKOS, A.P. and MARTIN, Q.W. (eds.). 1996. An international review of decision support systems in river basin operation. *Proceedings of the Fifth Water Resources Operations Management Workshop*, March Arlington, Va., ASCE.

GIJSBERS, P.J.A.; MOORE, R.V. and TINDALL, C.I. 2002. *Hamonil T: towards OMI, an open modelling interface and environment to harmonise European developments in water related simulation software*. Delft, the Netherlands, Delft Hydraulics (internal paper).

HAMLET, A. et al. 1996a. *Simulating basinwide alternatives using the ACT-ACF shared vision models*. Mobil District, Mobil, Al., USACE.

HAMLET, A. et al. 1996b. *A history of shared vision modelling in the ACT-ACF comprehensive study*. Mobil District, Mobil, Al., USACE.

HAMLET, A. et al. 1996c. *Basic STELLA II users manual for the ACT-ACF shared vision models*. Mobil District, Mobil, Al., USACE.

KINDLER, J. 1987. Systems analysis approach to water resources management: state of the art. *Symposium on recent developments and perspectives in systems analysis in water resources management*. Perugia, Italy, WARREDOC.

KINDLER, J. 1988. Systems analysis and water resources planning. *Fourth IFAC international symposium on systems analysis applied to management of water resources*. Rabat, Morocco, IFAC, 11 to 13 October.

LOUCKS, D.P. 1995. Developing and implementing decision support systems: a critique and a challenge. *Water Resources Bulletin, Journal of American Water Resources Association*, Vol. 31, No. 4, August, pp. 571–82.

LOUCKS, D.P. and DA COSTA, J.R. (eds.). 1991. *Decision support systems*. NATO Series G. Vol. 26. Berlin, Springer-Verlag.

LOUCKS, D.P.; STEDINGER, J.R. and SHAMIR, U. 1985. Modelling water resources systems: issues and experiences. *Civil Engineering Systems*, No. 2, pp. 223–31.

MAJONI, H. and QUADE, E.S. 1980. *Pitfalls of analysis*. New York, John Wiley.

MISER, H.J. 1980. Operations research and systems analysis. *Science*, No. 209, pp. 174–82.

O'CALLAGHAN, J.R. (ed.) 1996. The interaction of economic, ecology and hydrology, Chapman & Hall, London.

PALMER, R.N.; KEYS, A.M. and FISHER, S. 1995. Empowering stakeholders through simulation in water resources planning. *Proceedings of the 20th Conference on Water Resources Planning and Management Division*. New York, ASCE.

POOL, R. 1990. Struggling to do science for society. *Sci.*, No. 248, pp. 672–73.

RASKIN, P.; SIEBER, J. and HUBER-LEE, A. 2001. *Water evaluation and planning system: user guide for WEAP21*. Boston, Mass., Tellus Institute.

REYNOLDS, P.J. 1987. Future research needs in systems analysis and to application to water resources management. *International symposium on recent developments and perspectives in systems analysis in water resources management*. Perugia, Italy, WARREDOC.

ROGERS, P.P. and FIERING, M.B. 1986. Use of systems analysis in water management. *Water Resources Research*, Vol. 22, No. 9, pp. 146–58S.

SHAPIRO, H.T. 1990. The willingness to risk failure. *Science*, Vol. 250, No. 4981, p. 609.

SIMON, I-I.A. 1998. *Prediction and prescription in system modelling*. 15th Anniversary of IIASA, International Institute for Applied Systems Analysis. Laxenburg, Austria, IIASA.

STOKEY, E. and ZECKHAUSER, R. 1977. *A primer for policy analysis*. New York, W.W. Norton.

TOMLISON, R. 1980. Doing something about the future. *Journal of the Operational Research Society*, No. 31, pp. 467–76.

WALKER, W.E. 1987. *Important hard problems in public policy analysis*. Report P-7282. Santa Monica, Calif., Rand Corp.

WERICK, W.J.; WHIPPLE, W. Jr. and LUND, J. 1996. *ACT-ACF basinwide study*. Mobil District, Mobil, Al., USACE.

WHELAN, G. and NICHOLSON, T.J. (eds.). 2002. *Proceedings of the environmental software systems compability and linkage workshop, March 7–9, 2000*. NRC report NUREG/CP-0177. Richl and, Wash., Pacific Northwest National Laboratory.

# 3. Modelling Methods for Evaluating Alternatives

1. Introduction 59
   1.1. Model Components 60

2. Plan Formulation and Selection 61
   2.1. Plan Formulation 61
   2.2. Plan Selection 63

3. Modelling Methods: Simulation or Optimization 64
   3.1. A Simple Planning Example 65
   3.2. Simulation Modelling Approach 66
   3.3. Optimization Modelling Approach 66
   3.4. Simulation Versus Optimization 67
   3.5. Types of Models 69
      3.5.1. Types of Simulation Models 69
      3.5.2. Types of Optimization Models 70

4. Model Development 71

5. Managing Modelling Projects 72
   5.1. Create a Model Journal 72
   5.2. Initiate the Modelling Project 72
   5.3. Selecting the Model 73
   5.4. Analysing the Model 74
   5.5. Using the Model 74
   5.6. Interpreting Model Results 75
   5.7. Reporting Model Results 75

6. Issues of Scale 75
   6.1. Process Scale 75
   6.2. Information Scale 76
   6.3. Model Scale 76
   6.4. Sampling Scale 76
   6.5. Selecting the Right Scales 76

7. Conclusions 77

8. References 77

# 3 Modelling Methods for Evaluating Alternatives

Water resources systems are characterized by multiple interdependent components that together produce multiple economic, environmental, ecological and social impacts. Planners and managers working to improve the performance of these complex systems must identify and evaluate alternative designs and operating policies, comparing their predicted performance with the desired goals or objectives. These alternatives are defined by the values of numerous design, target and operating policy variables. Constrained optimization together with simulation modelling is the primary way we have of estimating the values of the decision variables that will best achieve specified performance objectives. This chapter introduces these optimization and simulation methods and describes what is involved in developing and applying them in engineering projects.

## 1. Introduction

Water resources system planners must identify and evaluate alternative water resources system designs or management plans on the basis of their economic, ecological, environmental, and social or political impacts. One important criterion for plan identification and evaluation is the economic benefit or cost a plan would entail were it to be implemented. Other criteria can include the extent to which any plan meets environmental, ecological and social targets. Once planning or management performance measures (objectives) and various general alternatives for achieving desired levels of these performance measures have been identified, models can be developed and used to help identify specific alternative plans that best meet those objectives.

Some system performance objectives may be in conflict, and in such cases models can help identify the efficient tradeoffs among these conflicting measures of system performance. These tradeoffs indicate what combinations of performance measure values can be obtained from various system design and operating policy variable values. If the objectives are the right ones (that is, they are what the stakeholders really care about), such quantitative tradeoff information should be of value during the debate over what decisions to make.

Regional water resources development plans designed to achieve various objectives typically involve investments in land and infrastructure. Achieving the desired economic, environmental, ecological and social objective values over time and space may require investments in storage facilities, including surface or groundwater reservoirs and storage tanks, pipes, canals, wells, pumps, treatment plants, levees and hydroelectric generating facilities, or in fact the removal of some of them.

Many capital investments can result in irreversible economic and ecological impacts. Once the forest of a valley is cleared and replaced by a lake behind a dam, it is almost impossible to restore the site to its original condition. In parts of the world where river basin or coastal restoration activities require the removal of engineering structures, water resources engineers are learning just how difficult and expensive that effort can be.

The use of planning models is not going to eliminate the possibility of making mistakes. These models can, however, better inform. They can provide estimates of the different impacts associated with, say, a natural unregulated river system and a regulated river system. The former can support a healthier ecosystem that provides a host of flood protection and water quality enhancement services. The latter can provide more reliable and cheaper water supplies for off-stream users and increased hydropower and some protection from at least small floods for those living on flood-prone lands. In short, models can help stakeholders assess the future consequences, the benefits and costs, and a multitude of other impacts associated with alternative plans or management policies.

This chapter introduces some mathematical optimization and simulation modelling approaches commonly used to study and analyse water resources systems. The modelling approaches are illustrated by their application to some relatively simple water resources planning and management problems. The purpose here is to introduce and compare some commonly used methods of (or approaches to) modelling. This is not a text on the state-of-the-art of optimization or simulation modelling. In subsequent chapters of this book, more details will be given about optimization models and simulation methods. More realistic and more complex problems usually require much bigger and more complex models than those developed in this book, but these bigger and more complex models are often based on the principles and techniques introduced here.

Regardless of the problem complexity or size, the modelling approaches are the same. Thus, the emphasis here is on the art of model development: just how one goes about constructing a model that will provide information needed to solve a particular problem, and various ways models might be solved. It is unlikely anyone will ever use any of the specific models developed in this or other chapters, simply because they will not be solving the specific examples used to illustrate the different approaches to model development and solution. However, it is quite likely that water resources managers and planners will use the modelling approaches and solution methods presented in this book to analyse similar types of problems. The particular problems used here, or any others that could have been used, can be the core of more complex models addressing more complex problems in practice.

Water resources planning and management today is dominated by the use of predictive optimization and simulation models. While computer software is becoming increasingly available for solving various types of optimization and simulation models, no software currently exists that will build those models themselves. What and what not to include and assume in models requires judgement, experience and knowledge of the particular problem being addressed, the system being modelled and the decision-making environment. Understanding the contents of, and performing the exercises for, this chapter will be a first step towards gaining some judgement and experience in model development.

## 1.1. Model Components

Mathematical models contain algebraic equations. These equations include variables that are assumed to be known and others that are unknown and to be determined. Known variables are usually called *parameters*, and unknown variables are called *decision variables*. Models are developed for the primary purpose of identifying the best values of the latter. These decision variables can include design and operating policy variables of various water resources system components.

Design variables can include the active and flood storage capacities of reservoirs, the power generating capacity of hydropower plants, the pumping capacity of pumping stations, the efficiencies of wastewater treatment plants, the dimensions or flow capacities of canals and pipes, the heights of levees, the hectares of an irrigation area, the targets for water supply allocations and so on. Operating variables can include releases of water from reservoirs or the allocations of water to various users over space and time. Unknown decision variables can also include measures of system performance, such as net economic benefits, concentrations of pollutants, ecological habitat suitability values or deviations from particular ecological, economic or hydrological targets.

Models describe, in mathematical terms, the system being analysed and the conditions that the system has to satisfy. These conditions are often called constraints. Consider, for example, a reservoir serving various water supply users downstream. The conditions included in a model of this reservoir would include the assumption that water will flow in the direction of lower heads (that is, downstream unless it is pumped upstream), and the volume of water stored in a reservoir cannot exceed the

**Figure 3.1.** These stakeholders have an interest in how their watershed or river basin is managed. Here they are using a physical model to help them visualize and address planning and management issues. Mathematical models often replace physical models, especially for planning and management studies. (Reprinted with permission from Engineering News-Record, copyright The McGraw-Hill Companies, Inc., September 20, 1993. All rights reserved.).

reservoir's storage capacity. Both the storage volume over time and the reservoir capacity might be unknown. If the capacity is known or assumed, then it is among the known model parameters.

Model parameter values, while assumed to be known, can often be uncertain. The relationships between various decision variables and assumed known model parameters (i.e., the model itself) may be uncertain. In these cases the models can be solved for a variety of assumed conditions and parameter values. This provides an estimate of just how important uncertain parameter values or uncertain model structures are with respect to the output of the model. This is called *sensitivity analysis*. Sensitivity analyses will be discussed in Chapter 9 in much more detail.

Solving a model means finding values of its unknown decision variables. The values of these decision variables can define a plan or policy. They can also define the costs and benefits or other measures of system performance associated with that particular management plan or policy.

While the components of optimization and simulation models include system performance indicators, model parameters and constraints, the process of model development and use includes people. The drawing shown in Figure 3.1 illustrates some interested stakeholders busy studying their river basin, in this case perhaps with the use of a physical model. Whether a mathematical model or physical model is being used, one important consideration is that if the modelling exercise is to be of any value, it must provide the information desired and in a form that the interested stakeholders can understand.

## 2. Plan Formulation and Selection

Plan formulation can be thought of as assigning particular values to each of the relevant decision variables. Plan selection is the process of evaluating alternative plans and selecting the one that best satisfies a particular objective or set of objectives. The processes of plan formulation and selection involve modelling and communication among all interested stakeholders, as the picture in Figure 3.1 suggests.

The planning and management issues being discussed by the stakeholders in the basin pictured in Figure 3.1 could well include surface and groundwater water allocations, reservoir operation, water quality management and infrastructure capacity expansion.

### 2.1. Plan Formulation

Model building for defining alternative plans or policies involves a number of steps. The first is to clearly specify the issue or problem or decision(s) to be made. What are the fundamental objectives and possible alternatives? Such alternatives might require defining allocations of water to various water users, the level of wastewater treatment, the capacities and operating rules of multipurpose reservoirs and hydropower plants, and the extent and reliability of floodplain protection from levees. Each of these decisions may affect system performance criteria or objectives. Often these objectives include economic measures of performance, such as costs and benefits. They may also include environmental and social measures not expressed in monetary units. (More detail on performance criteria is contained in Chapter 10.)

To illustrate this plan formulation process, consider the problem of designing a tank to hold a specific amount of water. The criterion to be used to compare different feasible designs is cost. The goal in this example is to find the least-cost shape and dimensions of a tank that will hold a specified volume, say $V$, of water.

The model of this problem must somehow relate the unknown design variable values to the cost of the tank. Assume, for example, a rectangular tank shape. The design variables are the length, $L$, width, $W$, and height, $H$, of the tank. These are the unknown decision variables. The objective is to find the combination of $L$, $W$, and $H$ values that minimizes the total cost of providing a tank capacity of at least $V$ units of water. This volume $V$ will be one of the model parameters. Its value is assumed known even though in fact it may be unknown and dependent in part on the cost.

The cost of the tank will be the sum of the costs of the base, the sides and the top. These costs will depend on the area of the base, sides and top. The costs per unit area may vary depending on the values of $L$, $W$ and $H$; however, even if those cost values depend on the values of those decision variables, given any specific values for $L$, $W$ and $H$, one can define an average cost-per-unit area. Here we will assume these average costs per unit area are known. They can be adjusted if they turn out to be incorrect for the derived values of $L$, $W$ and $H$.

These average unit costs of the base, sides and top will probably differ. They can be denoted as $C_{base}$, $C_{side}$ and $C_{top}$ respectively. These unit costs together with the tank's volume, $V$, are the parameters of the model. If $L$, $W$, and $H$ are measured in metres, then the areas will be expressed in units of square metres and the volume will be expressed in units of cubic metres. The average unit costs will be expressed in monetary units per square metre.

The final step of model building is to specify all the relations among the objective (cost), function and decision variables and parameters, including all the conditions that must be satisfied. It is often wise to first state these relationships in words. The result is a word model. Once that is written, mathematical notation can be defined and used to construct a mathematical model.

The word model for this tank design problem is to minimize total cost where:

- Total cost equals the sum of the costs of the base, the sides and the top.
- Cost of the sides is the cost-per-unit area of the sides times the total side area.
- Cost of the base is the cost-per-unit area of the base times the total base area.
- Cost of the top is the cost-per-unit area of the top times the total top area.
- The volume of the tank must at least equal some specified volume capacity.
- The volume of the tank is the product of the length, width and height of the tank.

Using the notation already defined, and combining some of the above conditions, a mathematical model can be written as:

$$\text{Minimize } Cost \tag{3.1}$$

Subject to:

$$Cost = (C_{base} + C_{top})(LW) + 2(C_{side})\,(LH + WH) \tag{3.2}$$

$$LWH \geq V \tag{3.3}$$

Equation 3.3 permits the tank's volume to be larger than that required. While this is allowed, it will cost more if the tank's capacity is larger than $V$, and hence the least-cost solution of this model will surely show that $LWH$ will equal $V$. In practice, however, there may be practical, legal and/or safety reasons why the decisions with respect to $L$, $W$ and $H$ may result in a capacity that exceeds the required volume, $V$.

This model can be solved a number of ways, which will be discussed later in this and the next chapters. The least-cost solution is

$$W = L = [2C_{side}\,V/(C_{base} + C_{top})]^{1/3} \tag{3.4}$$

$$\text{and} \quad H = V/[2C_{side}\,V/(C_{base} + C_{top})]^{2/3} \tag{3.5}$$

$$\text{or} \quad H = V^{1/3}[(C_{base} + C_{top})/2C_{side}]^{2/3} \tag{3.6}$$

The modelling exercise should not end here. If there is any doubt about the value of any of the parameters, a sensitivity analyses should be performed on those uncertain parameters or assumptions. In general these assumptions could include the values of the cost parameters (e.g., the costs-per-unit area) as well as the relationships expressed in the model (that is, the model itself). How much does the total cost change with respect to a change in any of the cost parameters or with the required volume $V$? How much does any decision-variable change with respect to changes in those parameter values? What is the percent change in a decision-variable value given a unit percent change in some parameter value (what economists call *elasticity*)?

If indeed the decision-variable values do not change significantly with respect to a change in the value of an uncertain parameter value, there is no need to devote more effort to reducing that uncertainty. Any time and money available for further study should be directed toward those parameters or assumptions that substantially influence the model's decision-variable values.

This capability of models to help identify what data are important and what data are not can guide monitoring and data collection efforts. This is a beneficial attribute of modelling often overlooked.

Continuing with the tank example, after determining, or estimating, all the values of the model parameters and then solving the model to obtain the cost-effective values of $L$, $W$ and $H$, we now have a design. It is just one of a number of designs that could be proposed. Another design might be for a cylindrical tank having a radius and height as decision-variables. For the same volume $V$ and unit area costs, we would find that the total cost is less, simply because the areas of the base, side and top are less. We could go one step further and consider the possibility of a truncated cone, having different bottom and top radii. In this case both radii and the height would be the decision-variables. But whatever the final outcome of our modelling efforts, there might be other considerations or criteria that are not expressed or included in the model that might be important to those responsible for plan (tank design) selection.

## 2.2. Plan Selection

Assume $P$ alternative plans (e.g., tank designs) have been defined, each designated by the index $p$. For each plan, there exist $n_p$ decision variables $x_j^p$ indexed with the letter $j$. Together these variables and their values, expressed by the vector $\boldsymbol{X}^p$, define the specifics of the $p$th plan. The index $j$ distinguishes one decision-variable from another, and the index $p$ distinguishes one plan from another. The task at hand, in this case, may be to find the particular plan $p$, defined by the known values of each decision-variable in the vector $\boldsymbol{X}^p$, that maximizes the present value of net benefits, $B(\boldsymbol{X}^p)$, derived from the plan.

Assume for now that an overall performance objective can be expressed mathematically as:

$$\text{maximize } B(\boldsymbol{X}^p) \tag{3.7}$$

The values of each decision-variable in the vector $\boldsymbol{X}^p$ that meet this objective must be feasible; in other words, they must meet all the physical, legal, social and institutional constraints.

$$\boldsymbol{X}^p \text{ feasible for all plans } p. \tag{3.8}$$

There are various approaches to finding the 'best' plan or best set of decision-variable values. By trial and error, one could identify alternative plans $p$, evaluate the net benefits derived from each plan, and select the particular plan whose net benefits are a maximum. This process could include a systematic simulation of a range of possible solutions in a search for the best. When there is a large number of feasible alternatives – that is, many decision-variables and many possible values for each of them – it may no longer be practical to identify and simulate all feasible combinations of decision-variable values, or even a small percentage of them. It would simply take too long. In this case it is often convenient to use an optimization procedure.

Equations 3.7 and 3.8 represent a discrete optimization problem. There are a finite set of discrete alternatives. The set could be large, but it is finite. The tank problem example is a continuous optimization problem having, at least mathematically, an infinite number of feasible solutions. In this case optimization involves finding feasible values of each decision-variable $x_j$ in the set of decision variables $\boldsymbol{X}$ that maximize (or minimize) some performance measure, $B(\boldsymbol{X})$. Again, feasible values are those that satisfy all the model constraints. A continuous constrained optimization problem can be written as:

$$\text{maximize } B(\boldsymbol{X}) \tag{3.9}$$

$$\boldsymbol{X} \text{ feasible} \tag{3.10}$$

While maximization of Equation 3.7 requires a comparison of $B(\boldsymbol{X}^p)$ for every discrete plan $p$, the maximization of Equation 3.9, subject to the feasibility conditions required in Equation 3.10, by complete enumeration is impossible. If there exists a feasible solution – in other words, at least one that satisfies all the constraints – mathematically there are likely to be an infinite number of possible feasible solutions or plans represented by various values of the decision-variables in the vector $\boldsymbol{X}$.

Finding by trial and error the values of the vector $\boldsymbol{X}$ that maximizes the objective Equation 3.9 and at the same time meet all the constraints is often difficult. Some type of optimization procedure, or algorithm, is useful in such cases. Mathematical optimization methods are designed to make this search for the best solution (or better solutions) more efficient. Optimization methods are used to identify those values of the decision-variables that satisfy specified objectives and constraints without requiring complete enumeration.

While optimization models might help identify the decision-variable values that will produce the best plan directly, they are based on all the assumptions incorporated in the model. Often these assumptions are limiting. In these cases the solutions resulting from optimization models should be analysed in more detail, perhaps through simulation methods, to improve the values of the decision-variables and to provide more accurate estimates of the impacts associated with those decision-variable values. In these situations, optimization models are used for screening out the clearly inferior solutions, not for finding the very best one. Just how screening is performed using optimization models will be discussed in the next chapter.

The values that the decision-variables may assume are rarely unrestricted. Usually various functional relationships among these variables must be satisfied. This is what is expressed in constraint Equations 3.8 and 3.10. For example, the tank had to contain a given amount of water. In a water-allocation problem, any water allocated to and completely consumed by one user cannot simultaneously or subsequently be allocated to another user. Storage reservoirs cannot store more water than their maximum capacity. Technological restrictions may limit the capacities and sizes of pipes, generators and pumps to those commercially available. Water quality concentrations should not exceed those specified by water quality standards or regulations. There may be limited funds available to spend on water resources development projects. These are a few examples of physical, legal and financial conditions or constraints that may restrict the ranges of variable values in the solution of a model.

Equations or inequalities can generally express any physical, economic, legal or social restrictions on the values of the decision-variables. Constraints can also simply define relationships among decision-variables. For example, Equation 3.2 above defines a new decision-variable called *Cost* as a function of other decision-variables and model parameters.

In general, constraints describe in mathematical terms the system being analysed. They define the system components and their inter-relationships, and the permissible ranges of values of the decision-variables, either directly or indirectly.

Typically, there exist many more decision-variables than constraints, and hence, if any feasible solution exists, there may be many such solutions that satisfy all the constraints. The existence of many feasible alternative plans is a characteristic of most water resources systems planning problems. Indeed it is a characteristic of most engineering design and operation problems. The particular feasible solution or plan that satisfies the objective function – that is, that maximizes or minimizes it – is called *optimal*. It is the optimal solution of the mathematical model, but it may not necessarily be considered optimal by any decision-maker. What is optimal with respect to some model may not be optimal with respect to those involved in a planning or decision-making process. To repeat what was written in Chapter 2, models are used to provide information (useful information, one hopes), to the decision-making process. Model solutions are not replacements for individuals involved in the decision-making process.

## 3. Modelling Methods: Simulation or Optimization

The modelling approach discussed in the previous section focused on the use of optimization methods to identify the preferred design of a tank. Similar methods can be used to identify preferred design-variable values and operating policies for multiple reservoir systems, for example. Once these preferred designs and operating policies have been identified, unless there is reason to believe that a particular alternative is really the best and needs no further analysis, each of these preferred alternatives can be further evaluated with the aid of more detailed and robust simulation models. Simulation models address *'what if'* questions: What will likely happen over time and at one or more specific places if a particular design and/or operating policy is implemented?

Simulation models are not limited by many of the assumptions incorporated into optimization models. For example, the inputs to simulation models can include a much longer time series of hydrological, economic and environmental data such as rainfall or streamflows, water supply demands, pollutant loadings and so on than would likely be included in an optimization model. The resulting outputs can better identify the variations of multiple system performance indicator values: that is, the multiple hydrological, ecological, economic and environmental impacts that might be observed over time, given any particular system design and operating policy.

Simulating multiple sets of values defining the designs and operating policies of a water resources system can take a long time. Consider, for example, only 30 decision-variables whose best values are to be determined. Even if only two values are assumed for each of the 30 variables, the number of combinations that could be simulated amounts to $2^{30}$ or in excess of $10^9$. Simulating and comparing even 1% of these billion at a minute per simulation amounts to over twenty years, continuously and without sleeping. Most simulation models of water resources systems contain many more variables and are much more complex than this simple 30-binary-variable example. In reality there could be an infinite combination of feasible values for each of the decision-variables.

Simulation works when there are only a relatively few alternatives to be evaluated, not when there are a large number of them. The trial and error process of simulation can be time consuming. An important role of optimization methods is to reduce the number of alternatives for simulation analyses. However, if only one method of analysis is to be used to evaluate a complex water resources system, simulation together with human judgement concerning which alternatives to simulate is often, and rightly so, the method of choice.

Simulation can be based on either discrete events or discrete time periods. Most simulation models of water resources systems are designed to simulate a sequence of discrete time periods. In each discrete time period, the simulation model converts all the initial conditions and inputs to outputs. The duration of each period depends in part on the particular system being simulated and the questions being addressed.

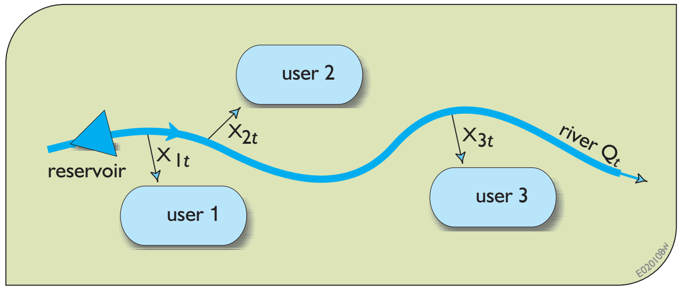



**Figure 3.2.** Reservoir-water allocation system to be simulated.

## 3.1. A Simple Planning Example

Consider the case of a potential reservoir releasing water to downstream users. A reservoir and its operating policy can increase the benefits each user receives over time by providing increased flows during periods of otherwise low flows relative to the user demands. Of interest is whether or not the increased benefits the water users obtain from an increased flow and more reliable downstream flow conditions will offset the costs of the reservoir. This water resources system is illustrated in Figure 3.2.

Before this system can be simulated, one has to define the active storage capacity of the reservoir and how much water is to be released depending on the storage volume and time period; in other words, one has to define the reservoir operating policy. In addition, one must also define the allocation policy: how much water to allocate to each user and to the river downstream of the users given any particular reservoir release.

For this simple illustration assume these operating and allocation policies are as shown in Figure 3.3. Also for simplicity assume they apply to each discrete time period.

The reservoir operating policy, shown as a red line in Figure 3.3, attempts to meet a release target. If insufficient water is available, all the water will be released in the time period. If the inflow exceeds the target flow and the reservoir is at capacity, a spill will occur. This operating policy is sometimes called the 'standard' operating policy. It is not usually followed in practice. Most operators, as indeed specified by most reservoir operating policies, will reduce releases in times of drought in an attempt to save

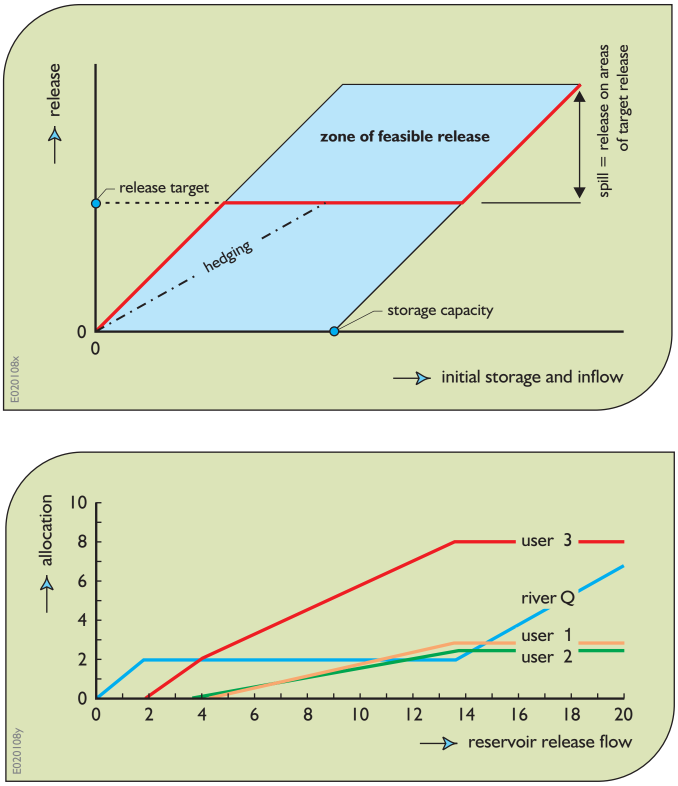



**Figure 3.3.** Policy defining the reservoir release to be made as a function of the current storage volume and current inflow and the reservoir release allocation policy for the river flow downstream of the reservoir. The blue zone in the reservoir release policy indicates the zone of feasible releases. It is physically impossible to make releases represented by points outside that blue zone.

some water in the reservoir for future releases in case of an extended period of low inflows. This is called a hedging policy. Any reservoir release policy, including a hedging policy, can be defined within the blue portion of the release policy plot shown in Figure 3.3. The dash–dot line in Figure 3.3 is one such hedging function.

Once defined, any reservoir operating policy can be simulated.

## 3.2. Simulation Modelling Approach

The simulation process for the three-user system shown in Figure 3.2 proceeds from one time period to the next. The reservoir inflow, obtained from a database, is added to the existing storage volume, and a release is determined from the release policy. Once the release is known, the final storage volume is computed and this becomes the initial volume for the next simulation time period. The reservoir release is then allocated to the three downstream users and to the river downstream of those users as defined by the allocation policy. The resulting benefits can be calculated and stored in an output database. Additional data pertaining to storage volumes, releases and the allocations themselves can also be stored in the output database, as desired. The process continues for the duration of the simulation run. Then the output data can be summarized for later comparison with other simulation results based on other reservoir capacities and operation policies and other allocation policies. Figure 3.4 illustrates this simulation process.

It would not be too difficult to write a computer program to carry out this simulation. In fact, it can be done on a spreadsheet. However easy that might be for anyone familiar with computer programming or spreadsheets, one cannot expect it to be easy for many practicing water resources planners and managers who are not doing this type of work on a regular basis. Yet they might wish to perform a simulation of their particular system, and to do it in a way that facilitates changes in many of its assumptions.

Computer programs capable of simulating a wide variety of water resources systems are becoming increasingly available. Simulation programs together with their interfaces that facilitate the input and editing of data and the display of output data are typically called decision support systems. Their input data define the components of the water resources system and their configuration. Inputs include hydrological data and design and operating policy data. These generic simulation programs are now becoming capable of simulating surface and groundwater water flows, storage volumes and qualities under a variety of system infrastructure designs and operating policies.

## 3.3. Optimization Modelling Approach

The simple reservoir-release and water-allocation planning example of Section 3.1 can also be described as an optimization model. The objective remains that of maximizing the total benefits that the three users obtain from the water that is allocated to them. Denoting each user's benefit as $B_{it}$ ($i = 1, 2, 3$) for each of $T$ time periods $t$, this objective, expressed symbolically is to:

$$\text{maximize total benefits} = \sum_{t}^{T} \{B_{1t} + B_{2t} + B_{3t}\} \tag{3.11}$$

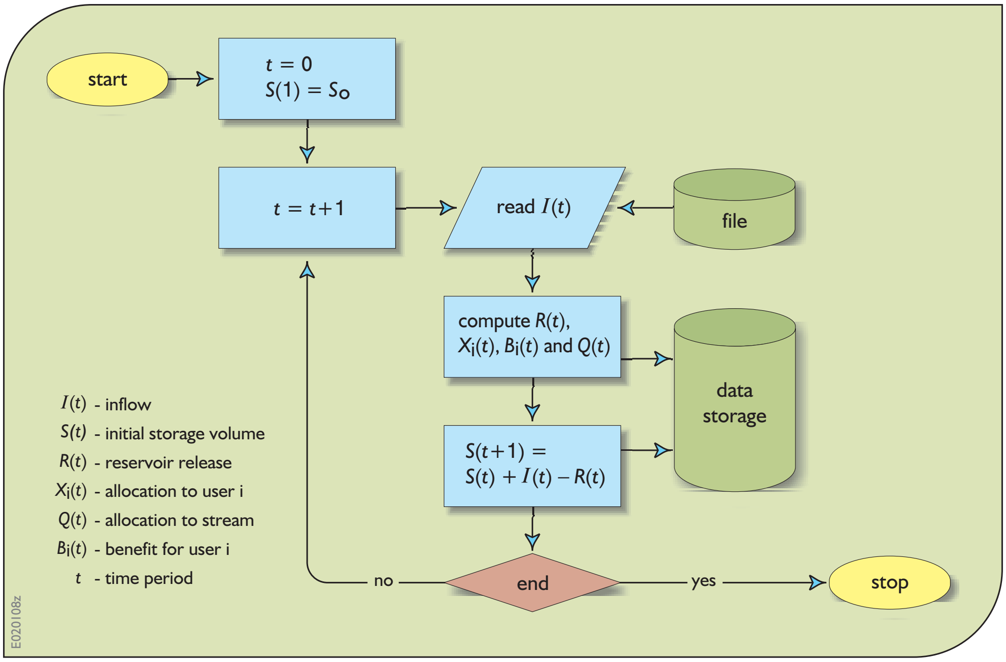



**Figure 3.4.** Flow diagram of the reservoir-allocation system simulation process. The simulation terminates after some predefined number of simulation time steps.

The benefit, $B_{it}$, for each user $i$ in each time period $t$ depends on the amount of water, $X_{it}$, allocated to it. These benefit functions, $B_{it} = B_{it}(X_{it})$, need to be known and expressed in a form suitable for solution using the particular optimization solution algorithm selected. The unknown variables include the allocations, $X_{it}$, and associated reservoir releases $R_t$ for all periods $t = 1, 2, 3, \dots, T$. Assuming there is no significant incremental runoff between the upstream reservoir and the sites where water is diverted from the river, the amounts allocated to all users, the sum of all $X_{it}$ in each period $t$, cannot exceed the amount of water released from the reservoir, $R_t$, in the period. This is one of the optimization model constraints:

$$R_t \geq X_{1t} + X_{2t} + X_{3t} \tag{3.12}$$

The remaining necessary constraints apply to the reservoir. A mass balance of water storage is needed, along with constraints limiting initial storage volumes, $S_t$, to the capacity, $K$, of the reservoir. Assuming a known time-series record of reservoir inflows, $I_t$, in each of the time periods being considered, the mass-balance or continuity equations for storage changes in each period $t$ can be written:

$$S_t + I_t - R_t = S_{t+1} \quad \text{for } t = 1, 2, \dots, T;$$
$$\text{If } t = T, \text{ then } T+1 = 1. \tag{3.13}$$

The capacity constraints simply limit the unknown initial storage volume, $S_t$, to be no greater than the reservoir capacity, $K$.

$$S_t \leq K \quad \text{for } t = 1, 2, \dots, T. \tag{3.14}$$

A one-year analysis period with $T = 12$ time periods of one month each in combination with three allocation variables, $X_{it}$, a storage variable $S_t$, and a release $R_t$, variable in each period $t$, includes a total of sixty unknown decision-variables. The job of the optimization solution procedure is to find the values of these sixty variables that will satisfy the objective, Equation 3.11, that is to say, maximize total benefits, and at the same time satisfy all of the thirty-six constraint equations and inequalities as well.

In this example the reservoir inflows, $I_t$, its storage capacity, $K$, and each user's benefit functions, $B_{it}$, are assumed known. In some cases such information is not known. Nor were other purposes, such as hydropower, flood control, water quality or recreation considered in this example, to mention only a few possible extensions. Such conditions and extensions will be considered in later chapters.

## 3.4. Simulation Versus Optimization

Unlike simulation models, the solutions of optimization models are based on objective functions of unknown decision-variables that are to be maximized or minimized.

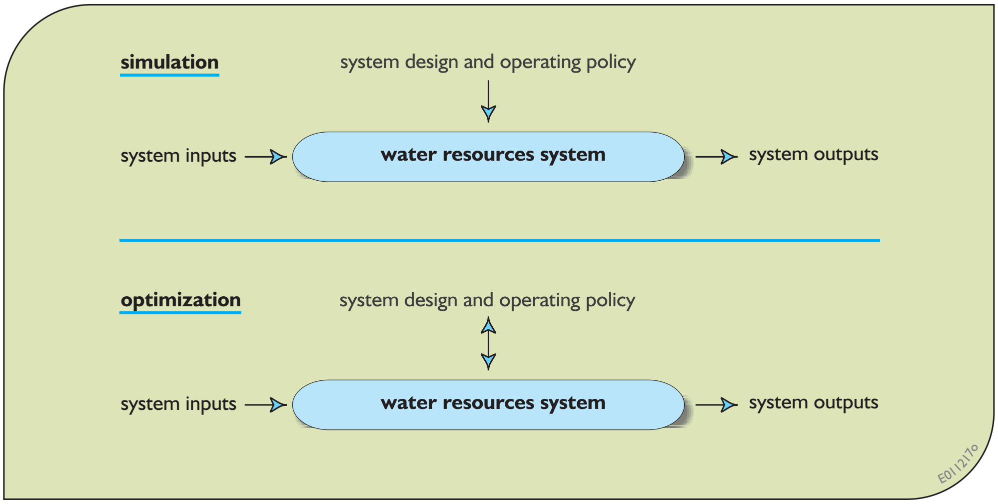



**Figure 3.5.** Distinguishing between simulation and optimization modelling. Simulation addresses '*what if*' questions; optimization can address '*what should be*' questions. Both types of models are often needed in water resources panning and management studies.

The constraints of an optimization model contain decision-variables that are unknown and parameters whose values are assumed known. Constraints are expressed as equations and inequalities. The tank model (Equations 3.1, 3.2 and 3.3) is an example of an optimization model. So is the reservoir water-allocation model, Equations 3.11 to 3.14. The solution of an optimization model, if one exists, contains the values of all of the unknown decision-variables. It is mathematically optimal in that the values of the decision-variables satisfy all the constraints and maximize or minimize an objective function. This 'optimal' solution is of course based on the assumed values of the model parameters, the chosen objective function and the structure of the model itself.

The procedure (or algorithm) most appropriate for solving any particular optimization model depends in part on the particular mathematical structure of the model. There is no single universal solution procedure that will efficiently solve all optimization models. Hence, model builders tend to model water resources systems by using mathematical expressions that are of a form compatible with one or more known solution procedures. Approximations of reality, made to permit model solution by a chosen optimization solution procedure (algorithm), may justify a more detailed simulation to check and improve on any solution obtained from that optimization. Simulation models are not restricted to any particular form of mathematics, and can define many relations including those not easily incorporated into optimization models.

One of many challenges in the use of optimization modelling is our inability to quantify and express mathematically all the planning objectives, the technical, economic, and political constraints and uncertainties, and other important considerations that will influence the decision-making process. Hence at best a mathematical-model of a complex water resources system is only an approximate description of the real system. The optimal solution of any model is optimal *only with respect to the particular model*, not necessarily with respect to the real system. It is important to realize this limited meaning of the word 'optimal,' a term commonly used by water resources and other systems analysts, planners and engineers.

Figure 3.5 illustrates the broad differences between simulation and optimization. Optimization models need explicit expressions of objectives. Simulation models do not. Simulation simply addresses 'what-if' sceanarios – what may happen if a particular scenario is assumed or if a particular decision is made. Users must specify the values of design and operating decision-variables before a simulation can be performed. Once these values of all decision-variables are defined, simulation can help us estimate more precisely the impacts that may result from those decisions. The difficulty with using simulation alone for analysing multiple alternatives occurs when there are many alternative, and potentially attractive, feasible solutions or plans and not enough time or resources to simulate them all. Even when combined with efficient techniques for selecting the values of each decision-variable, an enormous computational effort may still lead to a solution that is still far from the best possible.

For water resources planning and management, it is often advantageous to use both optimization and simulation modelling. While optimization will tell us

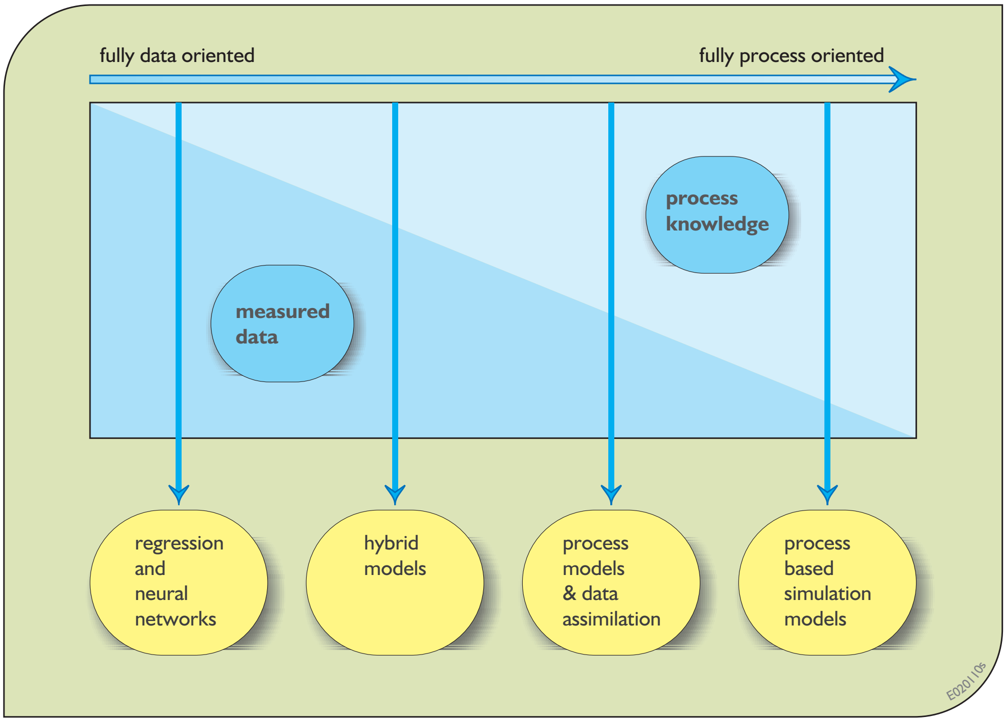



**Figure 3.6.** Range of simulation models types based on the extent to which measured field data and descriptions of system processes are included in the model.

what we should do – what the best decision is – that solution is often based on many limiting assumptions. Because of this, we need to use optimization not as a way to find the best solution, but to define a relatively small number of good alternatives that can later be tested, evaluated and improved by means of simulation. This process of using optimization to reduce the large number of plans and policies to a few that can then be simulated and better evaluated is often called preliminary screening.

## 3.5. Types of Models

### 3.5.1. Types of Simulation Models

Simulation models can be statistical or process oriented, or a mixture of both. Pure statistical models are based solely on data (field measurements). Pure process-oriented models are based on knowledge of the fundamental processes that are taking place. The example simulation model just discussed is a process-oriented model. It incorporated and simulated the physical processes taking place in the system. Many simulation models combine features of both of these extremes.

The range of various simulation modelling approaches applied to water resources systems is illustrated in Figure 3.6.

Regressions, such as that resulting from a least-squares analysis, and artificial neural networks are examples of purely statistical models. As discussed in Chapter 6, a relationship is derived between input data (cause) and output data (effect), based on measured and observed data. The relationship between the input and the output variable values is derived by calibrating a black-box or statistical model with a predefined structure unrelated to the actual natural processes taking place. Once calibrated, the model can be used to estimate the output variable values as long as the input variable values are within the range of those used to calibrate the model.

Hybrid models incorporate some process relationships into regression models or neural networks. These relationships supplement the knowledge contained in the calibrated parameter values based on measured data.

Most simulation models containing process relationships include parameters whose values need to be estimated. This is called model calibration. Calibration requires measured field data. These data can be used for initial calibration and verification, and in the case of ongoing simulations, for continual calibration and uncertainty reduction. The latter is sometimes referred to as data assimilation.

Other simulation model classifications are possible. Simulation models can be classified based on what the model simulates: on the domain of application. Today one can obtain, or develop, computer programs written to simulate a wide variety of water resources system components or events. Some of these include:

- water quantity and/or quality of rivers, bays, estuaries or coastal zones
- reservoir operation for quantity and/or quality
- saturated and/or unsaturated zone groundwater quantity and/or quality
- precipitation runoff, including erosion and chemicals
- water system demands, supply distribution and treatment
- high-water forecasting and control
- hazardous material spills
- morphological changes
- wastewater collection systems
- wastewater purification facilities
- irrigation operations
- hydropower production
- ecological habitats of wetlands, lakes, reservoirs and flood plains
- economic benefits and costs.

Simulation models of water resources systems generally have both spatial and temporal dimensions. These dimensions may be influenced by the numerical methods used, if any, in the simulation, but otherwise they are usually set, within the limits desired by the user. Spatial resolutions can range from 0 to 3 dimensions. Models are sometimes referred to as quasi 2- or 3-dimensional models. These are 1 or 2-dimensional models set up in a way that approximates what takes place in 2- or 3-dimensional space, respectively. A quasi-3D system of a reservoir could consist of a series of coupled 2D horizontal layers, for example.

Simulation models can be used to study what might occur during a given time period, say a year, sometime in the future, or what might occur from now to that given time in the future. Models that simulate some particular time in the future, where future conditions such as demands and infrastructure design and operation are fixed, are called stationary or static models. Models that simulate developments over time are called dynamic models. Static models are those in which the external environment of the system being simulated is not changing. Water demands, soil conditions, economic benefit and cost functions, populations and other factors do not change from one year to the next. Static models provide a snapshot or a picture at a point in time. Multiple years of input data may be simulated, but from the output statistical summaries can be made to identify what the values of all the impact variables could be, together with their probabilities, at that future time period.

Dynamic simulation models are those in which the external environment is also changing over time. Reservoir storage capacities could be decreasing due to sediment load deposition, costs could be increasing due to inflation, wastewater effluent discharges could be changing due to changes in populations and/or wastewater treatment capacities, and so on.

Simulation models can also vary in the way they are solved. Some use purely analytical methods while others require numerical ones. Many use both methods, as appropriate.

Finally, models can also be distinguished according to the questions being asked and the level of information desired. The type of information desired can range from data of interest to policy-makers and planners (requiring relatively simple models and broader in scope) to that of interest to researchers desiring a better understanding of the complex natural, economic and social processes taking place (requiring much more detailed models and narrower in scope). Water management and operational models (for real-time operations of structures, for example) and event-based calamity models are somewhere between these two extremes with respect to model detail. The scope and level of detail of any modelling study will also depend on the time, money and data available for that study (see Chapter 2).

### 3.5.2. Types of Optimization Models

There are many ways to classify various types of constrained optimization models. Optimization models can be deterministic or probabilistic, or a mixture of both. They can be static or dynamic with respect to time. Many water resources planning and management models are static, but include multiple time periods to obtain a statistical snapshot of various impacts in some planning period. Optimization models can be linear or non-linear. They can consist of continuous variables or discrete or integer variables, or a combination of both. But whatever type they are, they have in common the fact that they are describing situations where there exist multiple solutions that satisfy all the constraints, and hence, there is the desire to find the best solution, or at least a set of very good solutions.

Regardless of the type of optimization model, they all include an objective function. The objective function of an optimization model (such as Equation 3.11 in the example problem above) is used to evaluate multiple possible solutions. Often multiple objective functions may be identified (as will be discussed in Chapter 10). But at least one objective must be identified in all optimization models. Identifying the best objective function is often a challenging task.

Optimization models can be based on the particular type of application, such as reservoir sizing and/or operation, water quality management, or irrigation development or operation. Optimization models can also be classified into different types depending on the algorithm to be used to solve the model. Constrained optimization algorithms are numerous. Some guarantee to find the best model solution and others can only guarantee locally optimum solutions. Some include algebraic 'mathematical programming' methods and others include deterministic or random trial-and-error search techniques. Mathematical programming techniques include Lagrange multipliers, linear programming, non-linear programming, dynamic programming, quadratic programming, fractional programming and geometric programming, to mention a few. The applicability of each of these as well as other constrained optimization procedures is highly dependent on the mathematical structure of the model. The following Chapter 4 illustrates the application of some of the most commonly used constrained optimization techniques in water resources planning and management. These include classical constrained optimization using calculus-based Lagrange multipliers, discrete dynamic programming, and linear and non-linear programming.

Hybrid models usually include multiple solution methods. Many generic multi-period simulation models are driven by optimization methods within each time period. (The CALSIM II model used by the State of California and the US Bureau of Reclamation to allocate water in central California is one such model.)

Each of a variety of optimization modelling types and solution approaches will be discussed and illustrated in more detail in subsequent chapters. In some cases, we can use available computer programs to solve optimization models. In other cases, we may have to write our own software. To make effective use of optimization, and even simulation, models one has to learn some model solution methods, since those methods often dictate the type of model most appropriate for analysing a particular planning or management problem or issue.

To date, no single model type or solution procedure has been judged best for all the different types of issues and problems encountered in water resources planning and management. Each method has its advantages and limitations. One will experience these advantages and limitations as one practices the art of model development and application.

## 4. Model Development

Prior to the selection or development of a quantitative simulation model, it is often useful to develop a conceptual one. Conceptual models are non-quantitative representations of a system. The overall system structure is defined but not all its elements and functional relationships.

Figure 3.7 illustrates a conceptual model, without indicating what each box represents, defining relationships between what land and water managers can do and the eventual ecological impacts of those actions.

Once a conceptual model has been quantified (expressed in mathematical terms), it becomes a mathematical model. The model's equations typically include state and auxiliary variables, parameters and other model components.

The values of the model's parameters need to be determined. Model calibration involves finding the best values for these parameters. It is based on comparisons of the model results with field measurements. Optimization methods can sometimes be used to identify the values of all model parameters. This is called model calibration or identification. (Illustrations of the use of optimization for estimating model parameter values are contained in Chapters 4 and 6).

Sensitivity analysis (Chapter 9) may serve to identify the impacts of uncertain parameter values and show which parameter values substantially influence the

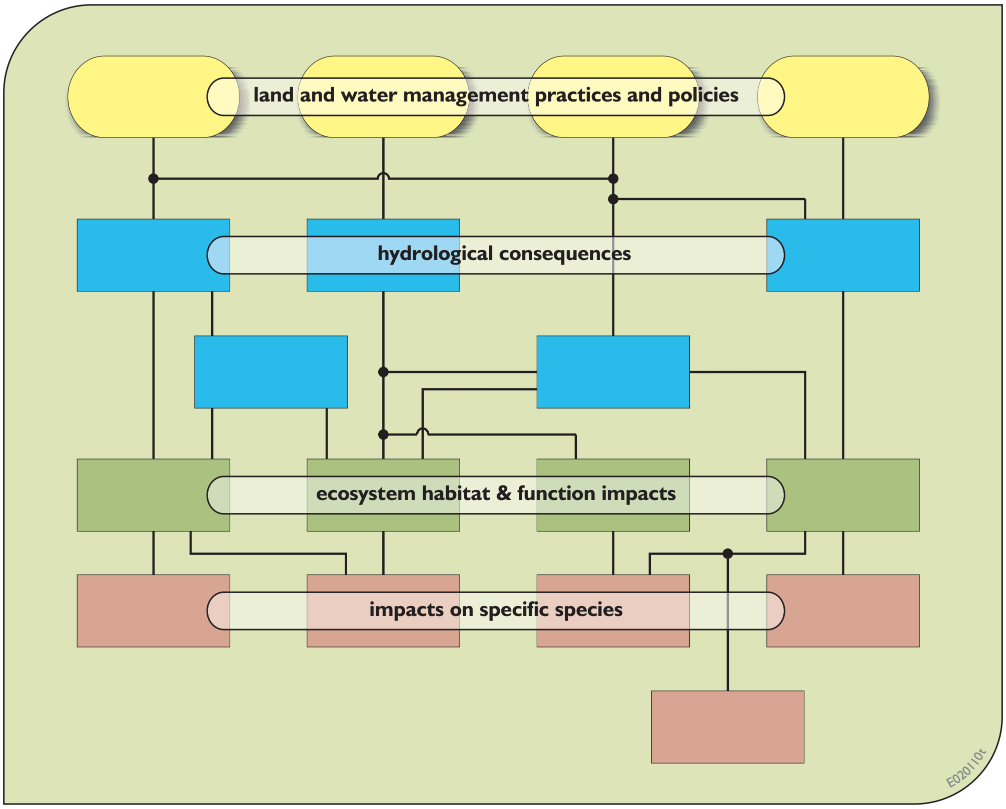



**Figure 3.7.** An example of a conceptual model without its detail, showing the cause and effect links between management decisions and specific system impacts.

model's results. Following calibration, the remaining uncertainties in the model predictions may be quantified in an uncertainty analysis as discussed in Chapter 9.

In addition to being calibrated, simulation models should also be validated or verified. In the validation or verification process the model results are compared with an independent set of measured observations that were not used in calibration. This comparison is made to verify whether or not the model describes the system behaviour sufficiently correctly.

# 5. Managing Modelling Projects

There are some steps that, if followed in modelling projects, can help reduce potential problems and lead to more effective outcomes. These steps are illustrated in Figure 3.8 (Scholten et al., 2000).

Some of the steps illustrated in Figure 3.8 may not be relevant in particular modelling projects. If so, these parts of the process can be skipped. Each of these modelling project steps is discussed in the next several sections.

## 5.1. Creating a Model Journal

One common problem of model studies once they are underway occurs when one wishes to go back over a series of simulation results to see what was changed, why a particular simulation was made or what was learned. It is also commonly difficult if not impossible for third parties to continue from the point at which any previous project terminated. These problems are caused by a lack of information on how the study was carried out. What was the pattern of thought that took place? Which actions and activities were carried out? Who carried out what work and why? What choices were made? How reliable are the end results? These questions should be answerable if a model journal is kept. Just like computer-programming documentation, this study documentation is often neglected under the pressure of time and perhaps because it is not as interesting as running the models themselves.

## 5.2. Initiating the Modelling Project

Project initiation involves defining the problem to be modelled and the objectives that are to be accomplished.

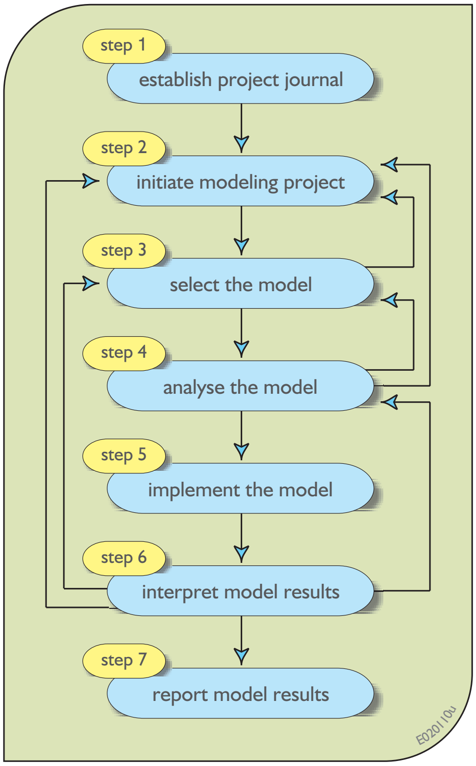



**Figure 3.8. The modelling project process is an iterative procedure involving specific steps or tasks.**

There can be major differences in perceptions between those who need information and those who are going to provide it. The problem 'as stated' is often not the problem 'as understood' by either the client or the modeller. In addition, problem perceptions and modelling objectives can change over the duration of a modelling project.

The appropriate spatial and time scales also need to be identified. The essential natural system processes must be identified and described. One should ask and answer the question of whether or not a particular modelling approach, or even modelling in general, is the best way to obtain the needed information. What are the alternatives to modelling or a particular modelling approach?

The objective of any modelling project should be clearly understood with respect to the domain and the problem area, the reason for using a particular model, the questions to be answered by the model, and the scenarios to be modelled. Throughout the project these objective components should be checked to see if any have changed and if they are being met.

The use of a model nearly always takes place within a broader context. The model itself can also be part of a larger whole, such as a network of models in which many are using the outputs of other models. These conditions may impose constraints on the modelling project.

Proposed modelling activities may have to be justified and agreements made where applicable. Any client at any time may wish for some justification of the modelling project activities. Agreement should be reached on how this justification will take place. Are intermediate reports required, have conditions been defined that will indicate an official completion of the modelling project, is verification by third parties required, and so on? It is particularly important to record beforehand the events or times when the client must approve the simulation results. Finally, it is also sensible to reach agreements with respect to quality requirements and how they are determined or defined, as well as the format, scope and contents of modelling project outputs (data files) and reports.

## 5.3. Selecting the Model

The selection of an existing model to be used in any project depends in part on the processes that will be modelled (perhaps as defined by the conceptual model), the data available and the data required by the model. The available data should include system observations for comparison of the model results. They should also include estimates of the degree of uncertainty associated with each of the model parameters. At a minimum this might only be estimates of the ranges of all uncertain parameter values. At best it could include statistical distributions of them. In this step of the process it is sufficient to know what data are available, their quality and completeness, and what to do about missing or outlier data.

Determining the boundaries of the model is an essential consideration in model selection. This defines what is to be included in a model and what is not. Any model selected will contain a number of assumptions. These assumptions should be identified and justified, and later tested.

Project-based matters such as the computers to be used, the available time and expertise, the modeller's personal preferences, and the client's wishes or requirements may also influence model choice. An important practical criterion is whether there is an accessible manual for operating the model program and a help desk available to address any possible problems.

The decision to use a model, and which model to use, is an important part of water resources plan formulation. Even though there are no clear rules on how to select the right model to use, a few simple guidelines can be stated:

- Define the problem and determine what information is needed and what questions need to be answered.
- Use the simplest method that will yield adequate accuracy and provide the answer to your questions.
- Select a model that fits the problem rather than trying to fit the problem to a model.
- Question whether increased accuracy is worth the increased effort and increased cost of data collection. (With the advances in computer technology, computational cost is rarely an issue except perhaps for some groundwater management problems.)
- Do not forget the assumptions underlying the model used and do not read more significance into the simulation results than is actually there.

## 5.4. Analysing the Model

Once a modelling approach or a particular model has been selected, its strengths and limitations should be studied in more detail. The first step is to set up a plan for testing and evaluating the model. These tests can include mass (and energy) balance checks and parameter sensitivity analyses (see Chapter 9). The model can be run under extreme input data conditions to see if the results are as expected.

Once a model is tested satisfactorily, it can be calibrated. Calibration focuses on the comparison between model results and field observations. An important principle is: the smaller the deviation between the calculated model results and the field observations, the better the model. This is indeed the case to a certain extent, as the deviations in a perfect model are only due to measurement errors. In practice, however, a good fit is by no means a guarantee of a good model.

The deviations between the model results and the field observations can be due to a number of factors. These include possible software errors, inappropriate modelling assumptions such as the (conscious) simplification of complex structures, neglect of certain processes, errors in the mathematical description or in the numerical method applied, inappropriate parameter values, errors in input data and boundary conditions, and measurement errors in the field observations.

To determine whether or not a calibrated model is 'good', it should be validated or verified. Calibrated models should be able to reproduce field observations not used in calibration. Validation can be carried out for calibrated models as long as an independent data set has been kept aside for this purpose. If all available data are used in the calibration process in order to arrive at the best possible results, validation will not be possible. The decision to leave out validation is often a justifiable one especially when data are limited.

Philosophically, it is impossible to know if a model of a complex system is 'correct'. There is no way to prove it. Experimenting with a model, by carrying out multiple validation tests, can increase one's confidence in that model. After a sufficient number of successful tests, one might be willing to state that the model is 'good enough', based on the modelling project requirements. The model can then be regarded as having been validated, at least for the ranges of input data and field observations used in the validation.

If model predictions are to be made for situations or conditions for which the model has been validated, one may have a degree of confidence in the reliability of those predictions. Yet one cannot be certain. Much less confidence can be placed on model predictions for conditions outside the range for which the model was validated.

While a model should not be used for extrapolations as commonly applied in predictions and in scenario analyses, this is often exactly the reason for the modelling project. What is likely to happen given events we have not yet experienced? A model's answer to this question should also include the uncertainties attached to these predictions. Beck (1987) summarized this dilemma in the following statement: 'using scientifically based models, you will often predict an incorrect future with great accuracy, and when using complex, non-identifiable models, you may be capable of predicting the correct future with great uncertainty'.

## 5.5. Using the Model

Once the model has been judged 'good enough', it may be used to obtain the information desired. One should develop a plan on how the model is to be used, identifying the input to be used, the time period(s) to be simulated, and the quality of the results to be expected. Again, close communication between the client and the modeller is essential, both in setting up this plan and throughout its implementation, to avoid any unnecessary misunderstandings about what information is wanted and the assumptions on which that information is to be based.

Before the end of this model-use step, one should determine whether all the necessary model runs have been performed and whether they have been performed well. Questions to ask include:

- Did the model fulfill its purpose?
- Are the results valid?
- Are the quality requirements met?
- Was the discretization of space and time chosen well?
- Was the choice of the model restrictions correct?
- Was the correct model and/or model program chosen?
- Was the numerical approach appropriate?
- Was the implementation performed correctly?
- Are the sensitive parameters (and other factors) clearly identified?
- Was an uncertainty analysis performed?

If any of the answers to these questions is no, then the situation should be corrected. If it cannot be corrected, then there should be a good reason for this.

### 5.6. Interpreting Model Results

Interpreting the information resulting from simulation models is a crucial step in a modelling project, especially in situations in which the client may only be interested in those results and not the way they were obtained. The model results can be compared to those of other similar studies. Any unanticipated results should be discussed and explained. The results should be judged with respect to the modelling project objectives.

The results of any water resources modelling project typically include large files of time-series data. Only the most dedicated of clients will want to read those files, so the data must be presented in a more concise form. Statistical summaries should explicitly include any restrictions and uncertainties in the results. They should identify any gaps in the domain knowledge, thus generating new research questions or identifying the need for more field observations and measurements.

### 5.7. Reporting Model Results

Although the results of a model should not be the sole basis for policy decisions, modellers have a responsibility to translate their model results into policy recommendations. Policy-makers, managers, and indeed the participating stakeholders often want simple, clear and unambiguous answers to complex questions. The executive summary of a report will typically omit much of the scientifically justified discussion in its main body regarding, say, the uncertainties associated with some of the data. This executive summary is often the only part read by those responsible for making decisions. Therefore, the conclusions of the model study must not only be scientifically correct and complete, but also concisely formulated, free of jargon, and fully understandable by managers and policy-makers. The report should provide a clear indication of the validity, usability and any restrictions of the model results. The use of visual aids, such as graphs and GIS, can be very helpful.

The final report should also include sufficient detail to allow others to reproduce the model study (including its results) and/or to proceed from the point where this study ended.

## 6. Issues of Scale

Scaling aspects play an important role in many modelling projects. Four different types of scales can be distinguished: the process scale, the information scale, the model scale and the sampling scale. Each of these is discussed below.

### 6.1. Process Scale

Most hydrological processes vary over space and time. The scale on which the process variations manifest themselves is referred to as the process scale. Spatially, process scale can vary from the movement of small granules of sediment, for example, to the flooding of large river basins or coastal zones. All kinds of intermediary scale processes can be found, such as drainage into ditches of runoff from parcels of land, transport of sediment in brooks and flow movements in aquifers.

Various temporal scales can also be distinguished, varying from the intensity of rain in less than a minute to the change in landscape in geological time. Many process descriptions require a spectrum of scales. Such is the case, for example, in the simulation of interdependent surface and groundwater quantity and quality processes taking place in a watershed.

## 6.2. Information Scale

Information scale is the spatial and temporal scale of the information required. Generally, a strategic water resources manager (for example the local, regional or national government) needs information on a scale relative to their responsibilities and authorities. This level of information is likely to differ from the level desired by operational water managers dealing with day-to-day issues.

Information at scales smaller than what is needed is seen as being 'noise'. Information at scales larger than what is needed is not relevant or helpful. For local organizations (e.g., water boards) concerned with runoff, for example, there is no need to collect information on individual raindrops. The important spatial variances are usually within a range varying from hundreds of metres to hundreds of kilometres. Larger-scale variations (differences between precipitation in the Netherlands and Russia or between North and South America, for example) are rarely if ever relevant. The information scale depends on the task set for the water planner or manager.

## 6.3. Model Scale

Model scales refer to their spatial and temporal discretization. The model scales determine the required data interpolation and aggregation.

If the temporal and spatial scales of the problem have not been defined clearly enough, this can affect the later phases of the modelling process negatively. If the model scale chosen is too large, this may result in too general a schematization and relevant details might not be derived from the results. If the chosen model scale is too small, irrelevant small-scale variations can lead to non-optimal calibrations for the large-scale variations.

In large, spatially distributed models in particular, it is vital that the scale and the number of independent parameters (degrees of freedom) are chosen on the basis of the available data. If too many parameters are included

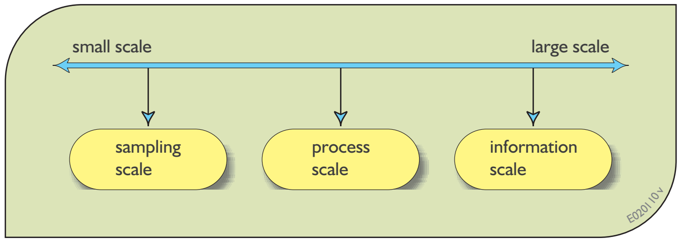



**Figure 3.9. Relationships among various scale types.**

in a model, there is a risk of it appearing to work well but being unsuitable for interpolation or prediction. This can actually only be determined if adequate measuring data are available having a measurement frequency at least equal to the chosen modelling time step. This must be taken into account when selecting or constructing the model, as there is otherwise the risk that the model cannot be calibrated adequately.

## 6.4. Sampling Scale

Sampling scale is the scale at which samples are taken. The sampling spatial scale can vary from 'point observations' (for example, a temperature measurement at a certain location at a certain time) to area observations (using, for example, remote sensing images). The density of the measuring network and the sampling or measuring and recording frequency determine the sampling spatial and temporal scales.

## 6.5. Selecting the Right Scales

Modellers must choose the model scale in such a manner that the model provides information at the required information scale, taking into account the process scales present, in combination with the spatial and temporal sampling scales. It is possible that situations will occur that are impossible to model just because of these scale issues.

The relationship between the types of scales is represented in Figure 3.9. The relative level of detail is given on the horizontal axis, from considerable detail on the left to much less detail on the right.

To show that the various types of scales may not be mutually compatible, consider the following three examples.

Example 1: *The information scale is different from the process scale*

Water resources planning studies are typically carried out at a river basin or watershed scale. From a hydrological (process) point of view, this makes sense because it enables a comprehensive analysis (including upstream and downstream impacts), makes it easier to develop a water balance for the study area, and reduces the amount of information needed at the borders of the study area. However, most decision-makers are not interested in results at river basin or watershed scale; they want to know what these results mean for their province, municipality or city. This conflict of scales can be solved by a well-considered selection of the (sub)watersheds that will be considered in the study and a post-processor that translates the results at these process scales into the required administrative scales.

Example 2: *The information scale is smaller than the process scale*

Imagine that a water resources manager wants to evaluate alternative anti-dehydration measures on the groundwater level over a period of five years. Thus the required information scale is a five-year period. However, the groundwater level is characterized by a very slow response. The relevant temporal groundwater-level process scale is around fifteen to twenty years. Whatever management alternative is implemented, it will take fifteen to twenty years to determine its impact. Thus, regardless of the choice of measuring frequency (sampling scale) and the model scale, it is impossible within the five-year period of interest to arrive at information on the groundwater-level changes as a consequence of anti-dehydration measures.

Example 3: *The sampling scale is larger than the process scale and information scale*

Assume it is necessary to estimate the change in concentrations of certain substances in the soil and groundwater in an urban area. The information spatial scale is one to two decimetres. This corresponds to the spatial variation of cohesion processes that take place in the soil and groundwater aquifer. However, logistic and budgetary considerations make it impossible to increase the spatial sampling density to less than a measurement site every few hundred metres. As the spatial sampling scale is much larger than the spatial process scale, useful interpolations cannot be made.

# 7. Conclusions

This chapter has reviewed some basic types of models and presented guidelines for consideration when undertaking a modelling project. Generic models for water resources system analyses are increasingly becoming available, saving many organizations that need model results from having to develop their own individual models. While many readers of this book may get involved in writing their own models, most of those involved in water resources planning and management will be using existing models and analysing and presenting their results. The information provided in this chapter is intended to help model users plan and manage effective modelling projects as well as improve the reproducibility, transferability and usefulness of model results.

# 8. References

BECK, M.B. 1987. Water quality modeling: a review of the analysis of uncertainty. *Water Resources Research*, No. 23, pp. 1393–1442.

SCHOLTEN, H.; WAVEREN, R.H. VAN; GROOT, S.; GEER, F.C. VAN; WÖSTEN, J.H.M.; KOEZE, R.D. and NOORT, J.J. 2000. *Good modelling practice in water management, proceedings hydroinformatics 2000*. Cedar Rapids, Iowa, International Association for Hydraulic Research.

## Additional References (Further Reading)

ABBOTT, M.B. 1999. Introducing hydroinformatics. *Journal of Hydroinformatics*, Vol. 1, No. 1, pp. 3–20.

BISWAS, A.K. (ed.). 1997. *Water resources: environmental planning, management, and development*. New York, McGraw-Hill.

BOBBA, A., GHOSH, S.; VIJAY, P. and BENGTSSON, L. 2000. Application of environmental models to different hydrological systems. *Ecological Modelling*, Vol. 125, pp. 15–49.

BOOCH, G. 1994. *Object-oriented analysis with applications*. Redwood City, Benjamin Cummings.

CHAN, H.W. and LAM, P.F. 1997. Visualizing Input and output analysis for simulation. *Simulation Practice and Theory*, Vol. 5, No. 5, pp. 425–53.

EDSALL, R.M.; HARROWER, M. and MENNIS, J.L. 2000. Tools for visualizing properties of spatial and temporal periodicities in geographic data. *Computers & Geosciences*, Vol. 26, No. 1, pp. 109–18.

FEDRA, K. and JAMIESON, D.G. 1996. An object-oriented approach to model integration: a river basin information system example. In: K. Kowar and H.P. Nachtnebel (eds.), *Proceedings of the HydroGIS'96 Conference, No. 235*, Vienna, IAHS Press. pp. 669–76.

GUARISO, G. and WERTHNER, H. 1989. *Environmental decision support systems*. New York, Ellis Horwood-Wiley.

HO, J.K.K. and SCULLI, D. 1997. The scientific approach to problem solving and decision support systems. *International Journal of Production Economics*, Vol. 48, No. 3, pp. 249–57.

HUFSCHMIDT, M.M. and FIERING, M.B. 1966. *Simulation in the design of water resource systems*. Cambridge, Mass., Harvard University Press.

LAM, D.C.L.; MAYFIELD, C.I. and SAWYNE, D.A. 1994. A prototype information system for watershed management and planning. *Journal of Biological Systems*, Vol. 2, No. 4, pp. 499–517.

LAM, D. and SWAYNE, D. 2001. Issues of EIS software design: some lessons learned in the past decade. *Environmental Modeling and Software*, Vol. 16, No. 5, pp. 419–25.

LAW, A.M. and KELTON, W.D. 2000. *Simulation modeling and analysis*, 3rd edn. Boston, Mass., McGraw-Hill.

LOUCKS, D.P.; STEDINGER, J.S. and HAITH, D.A. 1981. *Water resources systems planning and analysis*. Englewood Clifts, N.J., Prentice Hall.

MAASS, A.; HUFSCHMIDT, M.M.; DORFMAN, R.; THOMAS, H.A.; MARGLIN, S.A. and FAIR, G.M. 1966. *Design of water resources systems*. Cambridge, UK, Harvard University Press.

REITSMA, R.F. and CARRON, J.C. 1997. Object-oriented simulation and evaluation of river basin operations. *Journal of Geographic Information and Decision Analysis*, Vol. 1, No. 1, pp. 9–24.

RIZZOLI, A.E. and YOUNG, W.J. 1997. Delivering environmental decision support systems: software tools and techniques. *Environmental Modeling & Software*, Vol. 12, No. 2–3, pp. 237–49.

RIZZOLI, A.E.; DAVIS, J.R. and ABEL, D.J. 1998. Model and data integration and re-use in environmental decision support systems. *Decision support systems*, Vol. 24, No. 2, pp. 127–44.

SARGENT, R.G. 1982. Verification and validation of simulation models. In: F.E. Cellier (ed.), *Progress in modelling and simulation*, pp. 159–69. London and elsewhere, Academic Press.

SARGENT, R.G. 1984a. A tutorial on verification and validation of simulation models. In: S. Sheppard, U. Pooch and D. Pegden (eds.), *Proceedings of the 1984 winter simulation conference*, pp. 115–21. Piscataway, N.J., Institute of Electrical and Electronics Engineers.

SARGENT, R.G. 1984b. Simulation model validation. In: T.I. Ören, B.P. Zeigler and M.S. Elzas (eds.), *Simulation and model-based methodologies: an integrative view*, pp. 537–55. Berlin and elsewhere, Springer-Verlag (No. 10 in the series: NATO ASI Series F: Computer and Systems Sciences).

SINGH, V.P. (ed.). 1995. Computer models of watershed hydrology. Littleton, Colo., Water Resources Publications.

SUN, L. and LIU, K. 2001. A method for interactive articulation of information requirements for strategic decision support. *Information and Software Technology*, Vol. 43, No. 4, pp. 247–63.

VAN WAVEREN, R.H.; GROOT, S.; SCHOLTEN, H.; VAN GEER, F.C.; WÖSTEN, J.H.M.; KOEZE, R.D. and NOORT, J.J. 1999. *Good modelling practice handbook: STOWA report 99–05*. Lelystad, the Netherlands, Dutch Dept. of Public Works, Institute for Inland Water Management and Waste Water Treatment (report 99.036).

YOUNG, P. 1998. Data-based mechanistic modeling of environmental, ecological, economic and engineering systems. *Environmental Modeling & Software*, Vol. 13, pp. 105–22.

YOUNG, W.J.; LAM, D.C.L.; RESSEL, V. and WONG, J.W. 2000. Development of an environmental flows decision support system. *Environmental Modeling & Software*, Vol. 15, pp. 257–65.

ZEIGLER, B.P. 1976. *Theory of modelling and simulation*. Malabar, Fla, Robert E. Krieger.

ZEILER, M. 1999. *Modeling our world: the ESRI guide to geodatabase design*. Redlands, Calif., ESRI.

# 4. Optimization Methods


1. Introduction 81

2. Comparing Time Streams of Economic Benefits and Costs 81
   - 2.1. Interest Rates 82
   - 2.2. Equivalent Present Value 82
   - 2.3. Equivalent Annual Value 82

3. Non-Linear Optimization Models and Solution Procedures 83
   - 3.1. Solution Using Calculus 84
   - 3.2. Solution Using Hill Climbing 84
   - 3.3. Solution Using Lagrange Multipliers 86
     - 3.3.1. Approach 86
     - 3.3.2. Meaning of the Lagrange Multiplier 88

4. Dynamic Programming 90
   - 4.1. Dynamic Programming Networks and Recursive Equations 90
   - 4.2. Backward-Moving Solution Procedure 92
   - 4.3. Forward-Moving Solution Procedure 95
   - 4.4. Numerical Solutions 96
   - 4.5. Dimensionality 97
   - 4.6. Principle of Optimality 97
   - 4.7. Additional Applications 97
     - 4.7.1. Capacity Expansion 98
     - 4.7.2. Reservoir Operation 102
   - 4.8. General Comments on Dynamic Programming 112

5. Linear Programming 113
   - 5.1. Reservoir Storage Capacity–Yield Models 114
   - 5.2. A Water Quality Management Problem 117
     - 5.2.1. Model Calibration 118
     - 5.2.2. Management Model 119
   - 5.3. A Groundwater Supply Example 124
     - 5.3.1. A Simplified Model 125
     - 5.3.2. A More Detailed Model 126
     - 5.3.3. An Extended Model 127
     - 5.3.4. Piecewise Linearization Methods 128
   - 5.4. A Review of Linearization Methods 129

6. A Brief Review 132

7. Reference 132

# 4 Optimization Methods

Optimization methods are designed to provide the 'best' values of system design and operating policy variables – values that will lead to the highest levels of system performance. These methods, combined with more detailed and accurate simulation methods, are the primary ways we have, short of actually building physical models, of estimating the likely impacts of particular water resources system designs and operating policies. This chapter introduces and illustrates some of the art of optimization model development and use in analysing water resources systems. The modelling methods introduced in this chapter are extended in subsequent chapters.

## 1. Introduction

All of us are optimizers. We all make decisions that maximize our welfare in some way or another. Often the welfare we are maximizing may come later in life. By optimizing, it reflects our evaluation of future benefits versus current costs or benefits forgone. In economics, the extent to which we value future benefits today is reflected by what is called a *discount rate*. While economic criteria are only a part of everything we consider when making decisions, they are often among those deemed very important. Economic evaluation methods involving discount rates can be used to consider and compare alternatives characterized by various benefits and costs that occur over time. This chapter begins with a quick and basic review of what is often called *applied engineering economics*. Many of the optimization methods used in practice incorporate concepts from engineering economics.

Engineering economic methods typically identify a set of mutually exclusive alternatives (only one alternative can be selected) and then, using various methods involving the discount rate, identify the best one. The values of the decision-variables (e.g., the design and operating policy variables) are known for each alternative. For example, consider again the tank design problem presented in the previous chapter. Alternative tank designs could be identified, and then each could be evaluated, on the basis of cost and perhaps other criteria as well. The best would be called the optimal one, at least with respect to the objective criteria used.

The optimization methods introduced in this chapter extend those engineering economics methods. Some are discrete, some are continuous. Continuous optimization methods can identify the 'best' tank design, for example, without having to identify numerous discrete, mutually exclusive alternatives. Just how that can be done will be discussed later in this chapter.

Before proceeding to a more detailed discussion of optimization, a review of some methods of dealing with time streams of economic incomes or costs (engineering economics) may be useful.

## 2. Comparing Time Streams of Economic Benefits and Costs

Alternative plans, $p$, may involve different benefits and costs over time. These different plans need to be compared somehow. Let the net benefit generated in time period $t$ by plan $p$ be designated simply as $B_t^p$. Each plan is characterized by the time stream of net benefits it generates over its planning period $T_p$.

$$\{B_1^p, B_2^p, B_3^p, \ldots, B_{T_p}^p\} \tag{4.1}$$

Clearly, if in any time period $t$ the benefits exceed the costs, then $B_t^p > 0$; and if the costs exceed the benefits, $B_t^p < 0$. This section defines two ways of comparing different benefit, cost or net-benefit time streams produced by different plans perhaps having planning periods ending in different years $Tp$.

## 2.1. Interest Rates

Fundamental to the conversion of a time series of incomes and costs to a single value is the concept of the time value of money. From time to time, individuals, private corporations and governments need to borrow money to do what they want to do. The amount paid back to the lender has two components: (1) the amount borrowed and (2) an additional amount called *interest*. The interest amount is the cost of borrowing money, of having the money when it is loaned compared to when it is paid back. In the private sector the interest rate, the added fraction of the amount loaned that equals the interest, is often identified as the *marginal rate of return on capital*. Those who have money, called *capital*, can either use it themselves or they can lend it to others, including banks, and receive interest. Assuming people with capital invest their money where it yields the largest amount of interest, consistent with the risk they are willing to take, most investors should be receiving at least the prevailing interest rate as the return on their capital.

The interest rate includes a number of considerations. One is the *time value of money* (a willingness to pay something to obtain money now rather than to obtain the same amount later). Another is the *risk of losing capital* (not getting the full amount of a loan or investment returned at some future time). A third is the *risk of reduced purchasing capability* (the expected inflation over time). The greater the risks of losing capital or purchasing power, the higher the interest rate will be compared to the rate reflecting only the time value of money in a secure and inflation-free environment.

## 2.2. Equivalent Present Value

To compare projects or plans involving different time series of benefits and costs, it is often convenient to express these time series as a single equivalent value. One way to do this is to convert the time series to what it is worth today, its *present worth*, that is, a single value at the present time. This present worth will depend on the prevailing interest rate in each future time period. Assuming an amount $V_0$ is invested at the beginning of a time period, e.g., a year, in a project or a savings account earning interest at a rate $r$ per period, then at the end of the period the value of that investment is $(1 + r)V_0$.

If one invests an amount $V_0$ at the beginning of period $t = 1$ and at the end of that period immediately reinvests the total amount (the original investment plus interest earned), and continues to do this for $n$ consecutive periods at the same period interest rate $r$, the value, $V_n$, of that investment at the end of $n$ periods would be:

$$V_n = V_0(1 + r)^n \tag{4.2}$$

The initial amount $V_0$ is said to be equivalent to $V_n$ at the end of $n$ periods. Thus the present worth or present value, $V_0$, of an amount of money $V_n$ at the end of period $n$ is:

$$V_0 = V_n/(1 + r)^n \tag{4.3}$$

Equation 4.3 is the basic compound interest discounting relation needed to determine the present value at the beginning of period 1 of net benefits $V_n$ that accrue at the end of $n$ time periods.

The total present value of the net benefits generated by plan $p$, denoted $V_0^p$, is the sum of the values of the net benefits $V_t^p$ accrued at the end of each time period $t$ times the discount factor for that period $t$. Assuming the interest or discount rate $r$ in the discount factor applies for the duration of the planning period $Tp$,

$$V_0^p = \sum_t^{Tp} V_t^p/(1+r)^t \tag{4.4}$$

The present value of the net benefits achieved by two or more plans having the same economic planning horizons $Tp$ can be used as an economic basis for plan selection. If the economic lives or planning horizons of projects differ, then the present value of the plans may not be an appropriate measure for comparison and plan selection. A valid comparison of alternative plans is possible if all plans cover the same planning period.

## 2.3. Equivalent Annual Value

If the lives of various plans differ, but the same plans will be repeated on into the future, then one need only compare the equivalent constant annual net benefits of each plan. Finding the average or equivalent annual amount $V^p$ is done in two steps. First, one can compute the present value, $V_0^p$, of the time stream of net benefits, using Equation 4.4. The equivalent constant annual benefits, $V^p$, all discounted to the present must equal the present value, $V_0^p$.

$$V_0^p = \sum_t^{Tp} V_t^p/(1+r)^t = \sum_t^{Tp} V^p/(1+r)^t \tag{4.5}$$

With a little algebra the average annual end-of-year benefits $V^p$ of the project or plan $p$ is:

$$V^p = V_0^p[r(1 + r)^{Tp}]/[(1 + r)^{Tp} - 1] \tag{4.6}$$

The capital recovery factor $\text{CRF}_n$ is the expression $[r(1+r)^{Tp}]/[(1+r)^{Tp} - 1]$ in Equation 4.6 that converts a fixed payment or present value $V_0^p$ at the beginning of $n$ time periods into an equivalent fixed periodic payment $V^p$ at the end of each period. If the interest rate per period is $r$ and there are $n$ periods involved, then the capital recovery factor is:

$$\text{CRF}_n = [r(1 + r)^n]/[(1 + r)^n - 1] \tag{4.7}$$

This factor is often used to compute the equivalent annual end-of-year cost of engineering structures that have a fixed initial construction cost $C_0$ and annual end-of-year operation, maintenance, and repair (OMR) costs. The equivalent uniform end-of-year total annual cost, TAC, equals the initial cost times the capital recovery factor plus the equivalent annual end-of-year uniform OMR costs.

$$\text{TAC} = \text{CRF}_n C_0 + \text{OMR} \tag{4.8}$$

For private investments requiring borrowed capital, interest rates are usually established, and hence fixed, at the time of borrowing. However, benefits may be affected by changing interest rates, which are not easily predicted. It is common practice in benefit–cost analyses to assume constant interest rates over time, for lack of any better assumption.

Interest rates available to private investors or borrowers may not be the same rates that are used for analysing public investment decisions. In an economic evaluation of public-sector investments, the same relationships are used even though government agencies are not generally free to loan or borrow funds on private money markets. In the case of public-sector investments, the interest rate to be used in an economic analysis is a matter of public policy; it is the rate at which the government is willing to forego current benefits to its citizens in order to provide benefits to those living in future time periods. It can be viewed as the government's estimate of the time value of public monies or the marginal rate of return to be achieved by public investments.

These definitions and concepts of engineering economics are applicable to many of the problems faced in water resources planning and management. More detailed discussions of the application of engineering economics are contained in numerous texts on the subject.

## 3. Non-Linear Optimization Models and Solution Procedures

Constrained optimization is also called *mathematical programming*. Mathematical programming techniques include Lagrange multipliers, linear and non-linear programming, dynamic programming, quadratic programming, fractional programming and geometric programming, to mention a few. The applicability of each of these as well as other constrained optimization procedures is highly dependent on the mathematical structure of the model. The remainder of this chapter introduces and illustrates the application of some of the most commonly used constrained optimization techniques in water resources planning and management. These include classical constrained optimization using calculus-based Lagrange multipliers, discrete dynamic programming, and linear and non-linear programming.

Consider a river from which diversions are made to three water-consuming firms that belong to the same corporation, as illustrated in Figure 4.1. Each firm makes a product. Water is needed in the process of making that product, and is the critical resource. The three firms can be denoted by the index $j = 1$, 2 and 3 and their water allocations by $x_j$. Assume the problem is to determine the allocations $x_j$ of water to each of three firms ($j = 1, 2, 3$) that maximize the total net benefits, $\sum_j NB_j(x_j)$, obtained from all three firms. The total amount of water available is constrained or limited to a quantity of $Q$.

Assume the net benefits, $NB_j(x_j)$, derived from water $x_j$ allocated to each firm $j$, are defined by:

$$NB_1(x_1) = 6x_1 - x_1^2 \tag{4.9}$$

$$NB_2(x_2) = 7x_2 - 1.5x_2^2 \tag{4.10}$$

$$NB_3(x_3) = 8x_3 - 0.5x_3^2 \tag{4.11}$$

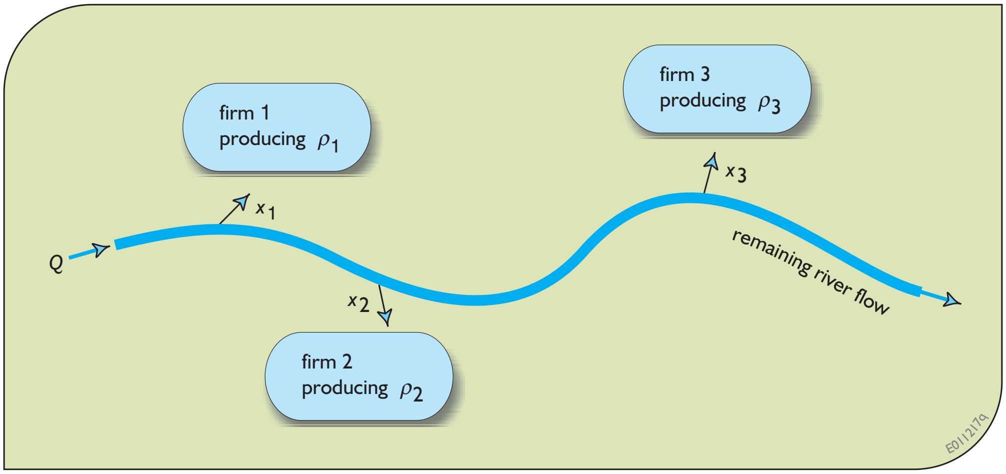



**Figure 4.1.** Three water-using firms obtain water from river diversions. The amounts allocated, $x_j$, to each firm $j$ will depend on the amount of water available, $Q$, in the river.

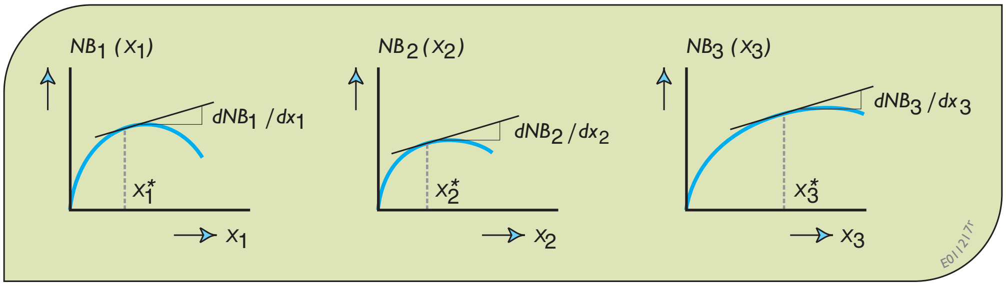



**Figure 4.2.** Concave net benefit functions and their slopes at allocations $x_1^*$, $x_2^*$ and $x_3^*$.

These are concave functions exhibiting decreasing marginal net benefits with increasing allocations. These functions look like hills, as illustrated in Figure 4.2.

## 3.1. Solution Using Calculus

Calculus can be used to find the allocation to each firm that maximizes its own net benefit, simply by finding where the slope or derivative of the net benefit function for the firm equals zero. The derivative, $\mathrm{d}NB(x_1)/\mathrm{d}x_1$, of the net benefit function for Firm 1 is $(6 - 2x_1)$ and hence the best allocation to Firm 1 would be 6/2 or 3. The best allocations for Firms 2 and 3 are 2.33 and 8 respectively. The total amount of water desired by all firms is the sum of each firm's desired allocation, or 13.33 flow units. However, suppose only 6 units of flow are available for all three firms. Introducing this constraint renders the previous solution infeasible. In this case we want to find the allocations that maximize the total net benefit obtained from all firms subject to having only 6 flow units to allocate. Using simple calculus will not suffice.

## 3.2. Solution Using Hill Climbing

One approach to finding the particular allocations that maximize the total net benefit derived from all firms in this example is an incremental steepest-hill-climbing method. This method divides the total available flow Q into increments and allocates each additional increment so as to get the maximum additional net benefit from that incremental amount of water. This procedure works in this example because the functions are concave; in other words, the marginal benefits decrease as the allocation increases. This procedure is illustrated by the flow diagram in Figure 4.3.

Table 4.1 lists the results of applying the procedure shown in Figure 4.3 to the problem of a) allocating 8 and b) allocating 20 flow units to the three firms and the river. Here a river flow of at least 2 is required and is to be satisfied, when possible, before any allocations are made to the firms.

The hill-climbing method illustrated in Figure 4.3 and Table 4.1 assigns each incremental flow $\Delta Q$ to the use that yields the largest additional (marginal) net benefit. An allocation is optimal for any total flow Q when the

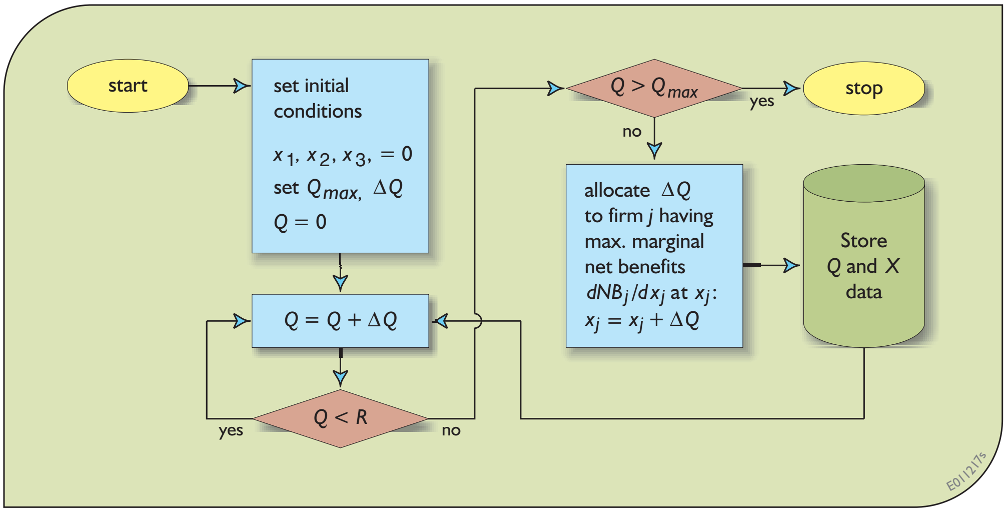



**Figure 4.3.** Steepest hill-climbing approach for finding allocation of a flow *Qmax* to the three firms, while meeting minimum river flow requirements *R*.

**Table 4.1.** Hill-climbing iterations for finding allocations that maximize total net benefit given a flow of *Qmax* and a required (minimum) streamflow of *R* = 2.

$Q_{max} = 8$; $Q_i = 0$; $\Delta Q = 1$; river flow $R \geq \min\{Q, 2\}$

| iteration $i$ | $Q_i$ | allocations. $R$ | $x_1$ | $x_2$ | $x_3$ | marginal net benefits $6\text{-}2x_1$ | $7\text{-}3x_2$ | $8\text{-}x_3$ | new allocations $Q_i + \Delta Q$ | $R$ | $x_1$ | $x_2$ | $x_3$ | total net benefits $\Sigma_j NB_a(x_1)_j$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-3 | 0-2 | 0-2 | 0 | 0 | 0 | 6 | 7 | 8 | 3 | 2 | 0 | 0 | 1 | 7.5 |
| 4 | 3 | 2 | 0 | 0 | 1 | 6 | 7 | 7 | 4 | 2 | 0 | 0 | 2 | 14.0 |
| 5 | 4 | 2 | 0 | 0 | 2 | 6 | 7 | 6 | 5 | 2 | 0 | 1 | 2 | 19.5 |
| 6 | 5 | 2 | 0 | 1 | 2 | 6 | 4 | 6 | 6 | 2 | 0 | 1 | 3 | 25.0 |
| 7 | 6 | 2 | 0 | 1 | 3 | 6 | 4 | 5 | 7 | 2 | 1 | 1 | 3 | 30.0 |
| 8 | 7 | 2 | 1 | 1 | 3 | 4 | 4 | 5 | 8 | 2 | 1 | 1 | 4 | 34.5 |
| 9 | 8 | 2 | 1 | 1 | 4 | 4 | 4 | 4 | – | – | – | – | – | ––– |

$Q_{max} = 20$; $\Delta Q \longrightarrow 0$; river flow $R \geq \min\{Q, 2\}$ selected values of $Q$

| $Q$ | allocations. $R$ | $x_1$ | $x_2$ | $x_3$ | marginal net benefits $6\text{-}2x_1$ | $7\text{-}3x_2$ | $8\text{-}x_3$ | total net benefits $\Sigma_j NB_j(x_1)$ |
|---|---|---|---|---|---|---|---|---|
| 0-2 | 0-2 | 0 | 0 | 0 | 6 | 7 | 8 | 0.0 |
| 4 | 2 | 0 | 0.25 | 1.75 | 6.00 | 6.25 | 6.25 | 14.1 |
| 5 | 2 | 0.18 | 0.46 | 2.36 | 5.64 | 5.64 | 5.64 | 20.0 |
| 8 | 2 | 1.00 | 1.00 | 4.00 | 4.00 | 4.00 | 4.00 | 34.5 |
| 10 | 2 | 1.55 | 1.36 | 5.09 | 2.91 | 2.91 | 2.91 | 41.4 |
| 15 | 2 | 2.91 | 2.27 | 7.82 | 0.18 | 0.18 | 0.18 | 49.1 |
| 20 | 6.67 | 3.00 | 2.33 | 8.00 | 0 | 0 | 0 | 49.2 |

E020820g

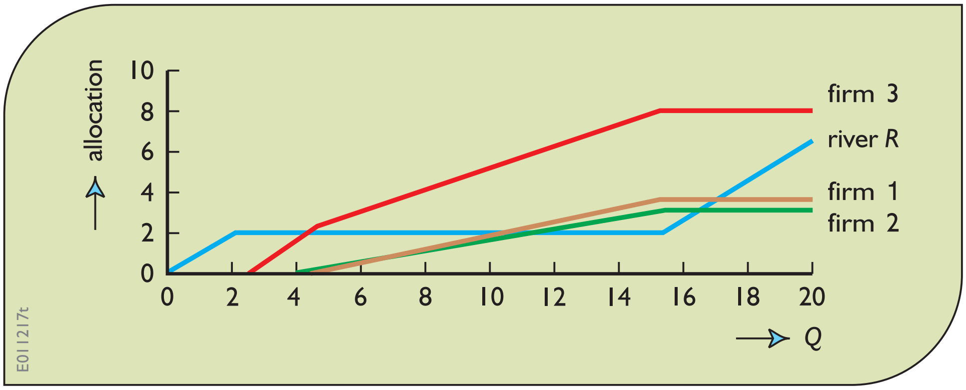



**Figure 4.4.** Water allocation policy that maximizes total net benefits derived from all three water-using firms.

marginal net benefits from each non-zero allocation are equal, or as close to each other as possible given the size of the increment $\Delta Q$. In this example, with a $\Delta Q$ of 1 and Qmax of 8, it just happens that the marginal net benefits are all equal (to 4). The smaller the $\Delta Q$, the more optimal will be the allocations in each iteration, as shown in the lower portion of Table 4.1 where $\Delta Q$ approaches 0.

Based on the allocations derived for various values of available water $Q$, as shown in Table 4.1, an allocation policy can be defined. For this problem the allocation policy that maximizes total net benefits is shown in Figure 4.4.

This hill-climbing approach leads to optimal allocations only if all of the net benefit functions whose sum is being maximized are concave: that is, the marginal net benefits decrease as the allocation increases. Otherwise, only a local optimum solution can be guaranteed. This is true using any calculus-based optimization procedure or algorithm.

## 3.3. Solution Using Lagrange Multipliers

### 3.3.1. Approach

As an alternative to hill-climbing methods, consider a calculus-based method involving Lagrange multipliers. To illustrate this approach, a slightly more complex water-allocation example will be used. Assume that the benefit, $B_j(x_j)$, each water-using firm receives is determined, in part, by the quantity of product it produces and the price per unit of the product that is charged. As before, these products require water and water is the limiting resource. The amount of product produced, $p_j$, by each firm $j$ is dependent on the amount of water, $x_j$, allocated to it.

Let the function $P_j(x_j)$ represent the maximum amount of product, $p_j$, that can be produced by firm $j$ from an allocation of water $x_j$. These are called production functions. They are typically concave: as $x_j$ increases the slope, $dP_j(x_j)/dx_j$, of the production function, $P_j(x_j)$, decreases.

For this example assume the production functions for the three water-using firms are:

$$p_1 = 0.4(x_1)^{0.9} \quad (4.12)$$

$$p_2 = 0.5(x_2)^{0.8} \quad (4.13)$$

$$p_3 = 0.6(x_3)^{0.7} \quad (4.14)$$

Next consider the cost of production. Assume the associated cost of production can be expressed by the following convex functions:

$$C_1 = 3(p_1)^{1.3} \quad (4.15)$$

$$C_2 = 5(p_2)^{1.2} \quad (4.16)$$

$$C_3 = 6(p_3)^{1.15} \quad (4.17)$$

Each firm produces a unique patented product, and hence it can set and control the unit price of its product. The lower the unit price, the greater the demand and thus the more each firm can sell. Each firm has determined the relationship between the amount that can be sold and the unit price – that is, the demand functions for that product. These unit price or demand functions are shown in Figure 4.5 where the $p_j$'s are the amounts of each product produced. The vertical axis of each graph is the unit price. To simplify the problem we are assuming linear demand functions, but this assumption is not a necessary condition.

The optimization problem is to find the water allocations, the production levels and the unit prices that together maximize the total net benefit obtained from all

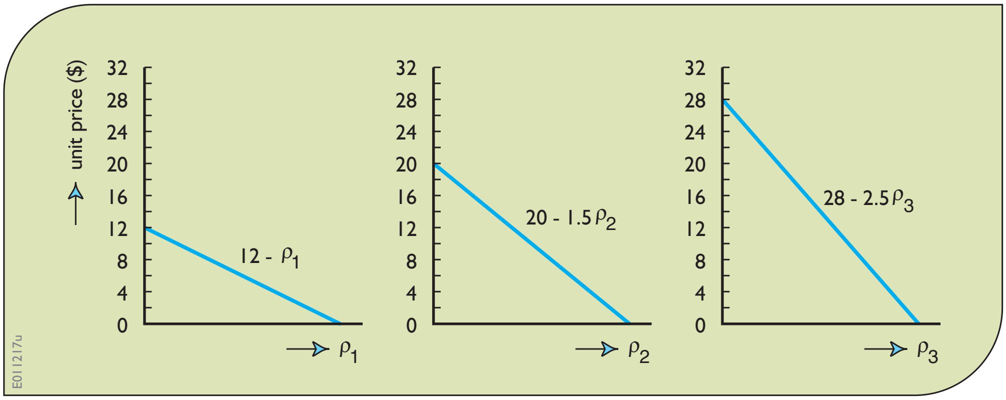



**Figure 4.5.** Unit prices that will guarantee the sale of the specified amounts of products $p_j$ produced in each of the three firms.

three firms. The water allocations plus the amount that must remain in the river, $R$, cannot exceed the total amount of water $Q$ available.

Constructing and solving a model of this problem for various values of $Q$, the total amount of water available, will define the three allocation policies as functions of $Q$. These policies can be displayed as a graph, as in Figure 4.4, showing the three best allocations given any value of $Q$. This of course assumes the firms can adjust to varying values of $Q$. In reality this may not be the case. (Chapter 10 examines this problem using more realistic benefit functions that reflect the degree to which firms can adapt to changing inputs over time.)

The model:

$$\text{Maximize Net\_benefit} \tag{4.18}$$

Subject to:

Definitional constraints:

$$\text{Net\_benefit} = \text{Total\_return} - \text{Total\_cost} \tag{4.19}$$

$$\text{Total\_return} = (12 - p_1)p_1 + (20 - 1.5p_2)p_2 + (28 - 2.5p_3)p_3 \tag{4.20}$$

$$\text{Total\_cost} = 3(p_1)^{1.30} + 5(p_2)^{1.20} + 6(p_3)^{1.15} \tag{4.21}$$

Production functions defining the relationship between water allocations $x_j$ and production $p_j$:

$$p_1 \le 0.4(x_1)^{0.9} \tag{4.22}$$

$$p_2 \le 0.5(x_2)^{0.8} \tag{4.23}$$

$$p_3 \le 0.6(x_3)^{0.7} \tag{4.24}$$

Water allocation restriction:

$$R + x_1 + x_2 + x_3 \le Q \tag{4.25}$$

One can first solve this model for the values of each $p_j$ that maximize the total net benefits, assuming water is not a limiting constraint. This is equivalent to finding each individual firm's maximum net benefits, assuming all the water that is needed is available. Using calculus we can equate the derivatives of the total net benefit function with respect to each $p_j$ to 0 and solve each of the resulting three independent equations:

$$\text{Total Net\_benefit} = [(12 - p_1)p_1 + (20 - 1.5p_2)p_2 + (28 - 2.5p_3)p_3] - [3(p_1)^{1.30} + 5(p_2)^{1.20} + 6(p_3)^{1.15}] \tag{4.26}$$

Derivatives:

$$\partial(\text{Net\_benefit})/\partial p_1 = 0 = 12 - 2p_1 - 1.3(3)p_1^{0.3} \tag{4.27}$$

$$\partial(\text{Net\_benefit})/\partial p_2 = 0 = 20 - 3p_2 - 1.2(5)p_2^{0.2} \tag{4.28}$$

$$\partial(\text{Net\_benefit})/\partial p_3 = 0 = 28 - 5p_3 - 1.15(6)p_3^{0.15} \tag{4.29}$$

The result (rounded off) is $p_1 = 3.2$, $p_2 = 4.0$, and $p_3 = 3.9$ to be sold for unit prices of 8.77, 13.96, and 18.23 respectively, for a maximum net revenue of 155.75. This would require water allocations $x_1 = 10.2$, $x_2 = 13.6$ and $x_3 = 14.5$, totaling 38.3 flow units. Any amount of water less than 38.3 will restrict the allocation to, and hence the production at, one or more of the three firms.

If the total available amount of water is less than that desired, constraint Equation 4.25 can be written as an equality, since all the water available, less any that must remain in the river, $R$, will be allocated. If the available water supplies are less than the desired 38.3 plus the required streamflow $R$, then Equations 4.22 through 4.25 need to be added. These can be rewritten as equalities since they will be binding.

$$p_1 - 0.4(x_1)^{0.9} = 0 \tag{4.30}$$

$$p_2 - 0.5(x_2)^{0.8} = 0 \tag{4.31}$$

$$p_3 - 0.6(x_3)^{0.7} = 0 \tag{4.32}$$

$$R + x_1 + x_2 + x_3 - Q = 0 \tag{4.33}$$

The first three constraints, Equations 4.30, 4.31 and 4.32, are the production functions specifying the relationships between each water input $x_j$ and product output $p_j$. The fourth constraint Equation 4.33 is the restriction on the total allocation of water. Since each of the four constraint equations equals zero, each can be added to the net benefit Equation 4.26 without changing its value. This is done in Equation 4.34, which is equivalent to Equation 4.26. The variable $L$ is the value of the Lagrange form of the objective function:

$$\begin{aligned} L = {} & [(12 - p_1)p_1 + (20 - 1.5p_2)p_2 + (28 - 2.5p_3)p_3] \\ & - [3(p_1)^{1.30} + 5(p_2)^{1.20} + 6(p_3)^{1.15}] \\ & - \lambda_1[p_1 - 0.4(x_1)^{0.9}] - \lambda_2[p_2 - 0.5(x_2)^{0.8}] \\ & - \lambda_3[p_3 - 0.6(x_3)^{0.7}] - \lambda_4[R + x_1 + x_2 + x_3 - Q] \end{aligned} \tag{4.34}$$

Since each of the four constraint Equations 4.30 through 4.33 included in Equation 4.34 equals zero, each can be multiplied by a variable $\lambda_i$ without changing the value of Equation 4.34. These unknown variables $\lambda_i$ are called the Lagrange multipliers of constraints $i$. The value of each multiplier, $\lambda_i$, is the marginal value of the original objective function, Equation 4.26, with respect to a change in the value of the amount produced, $p_i$, or in the case of constraint, Equation 4.33, the amount of water available, $Q$. We will prove this shortly.

Differentiating Equation 4.34 with respect to each of the ten unknowns and setting the resulting equations to 0 yields:

$$\partial L/\partial p_1 = 0 = 12 - 2p_1 - 1.3(3)p_1^{0.3} - \lambda_1 \tag{4.35}$$

$$\partial L/\partial p_2 = 0 = 20 - 3p_2 - 1.2(5)p_2^{0.2} - \lambda_2 \tag{4.36}$$

$$\partial L/\partial p_3 = 0 = 28 - 5p_3 - 1.15(6)p_3^{0.15} - \lambda_3 \tag{4.37}$$

$$\partial L/\partial x_1 = 0 = \lambda_1 0.9(0.4)(x_1)^{-0.1} - \lambda_4 \tag{4.38}$$

$$\partial L/\partial x_2 = 0 = \lambda_2 0.8(0.5)(x_2)^{-0.2} - \lambda_4 \tag{4.39}$$

$$\partial L/\partial x_3 = 0 = \lambda_3 0.7(0.6)(x_3)^{-0.3} - \lambda_4 \tag{4.40}$$

$$\partial L/\partial \lambda_1 = 0 = p_1 - 0.4(x_1)^{0.9} \tag{4.41}$$

$$\partial L/\partial \lambda_2 = 0 = p_2 - 0.5(x_2)^{0.8} \tag{4.42}$$

$$\partial L/\partial \lambda_3 = 0 = p_3 - 0.6(x_3)^{0.7} \tag{4.43}$$

$$\partial L/\partial \lambda_4 = 0 = R + x_1 + x_2 + x_3 - Q \tag{4.44}$$

These ten equations are the conditions necessary for an optimal solution. They can be solved to obtain the values of the ten unknown variables. The solutions to these equations for various values of $Q$, (found in this case by using LINGO) are shown in Table 4.2. (A demo version of LINGO is available, together with its help files, at www.lindo.com.)

### 3.3.2. Meaning of the Lagrange Multiplier

In this example, Equation 4.34 is the objective function. It is maximized (or minimized) by equating to zero each

**Table 4.2. Solutions to Equations 4.35 through 4.44.**

| water | allocations to firms | | | product productions | | | Lagrange multipliers | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| available | 1 | 2 | 3 | 1 | 2 | 3 | marginal net benefits | | | |
| Q-R | $x_1$ | $x_2$ | $x_3$ | $p_1$ | $p_2$ | $p_3$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ |
| 10 | 1.2 | 3.7 | 5.1 | 0.46 | 1.44 | 1.88 | 8.0 | 9.2 | 11.0 | 2.8 |
| 20 | 4.2 | 7.3 | 8.5 | 1.46 | 2.45 | 2.68 | 4.7 | 5.5 | 6.6 | 1.5 |
| 30 | 7.5 | 10.7 | 11.7 | 2.46 | 3.34 | 3.37 | 2.0 | 2.3 | 2.9 | 0.6 |
| 38 | 10.1 | 13.5 | 14.4 | 3.20 | 4.00 | 3.89 | 0.1 | 0.1 | 0.1 | 0.0 |
| 38.3 | 10.2 | 13.6 | 14.5 | 3.22 | 4.02 | 3.91 | 0 | 0 | 0 | 0 |

E020820h

of its partial derivatives with respect to each unknown variable. Equation 4.34 consists of the original net benefit function plus each constraint $i$ multiplied by a weight or multiplier $\lambda_i$. This equation is expressed in monetary units, such as dollars or euros. The added constraints are expressed in other units: either the quantity of product produced or the amount of water available. Thus the units of the weights or multipliers $\lambda_i$ associated with these constraints are expressed in monetary units per constraint units. In this example the multipliers $\lambda_1$, $\lambda_2$ and $\lambda_3$ represent the change in the total net benefit value of the objective function (Equation 4.26) per unit change in the products $p_1$, $p_2$ and $p_3$ produced. The multiplier $\lambda_4$ represents the change in the total net benefit per unit change in the water available for allocation, $Q - R$.

Note in Table 4.2 that as the quantity of available water increases, the marginal net benefits decrease. This is reflected in the values of each of the multipliers, $\lambda_i$. In other words, the net revenue derived from a quantity of product produced at each of the three firms, and from the quantity of water available, are concave functions of those quantities, as illustrated in Figure 4.2.

To review the general Lagrange multiplier approach and derive the definition of the multipliers, consider the general constrained optimization problem containing $n$ decision-variables $x_j$ and $m$ constraint equations $i$.

$$\text{Maximize (or minimize) } F(\boldsymbol{X}) \tag{4.45}$$

subject to constraints

$$g_i(\boldsymbol{X}) = b_i \quad i = 1, 2, 3, \ldots, m \tag{4.46}$$

where $\boldsymbol{X}$ is the vector of all $x_j$. The Lagrange function $L(\boldsymbol{X}, \boldsymbol{\lambda})$ is formed by combining Equations 4.46, each equalling zero, with the objective function of Equation 4.45.

$$L(\boldsymbol{X}, \boldsymbol{\lambda}) = F(\boldsymbol{X}) - \sum_i \lambda_i (g_i(\boldsymbol{X}) - b_i) \tag{4.47}$$

Solutions of the equations

$\partial L/\partial x_j = 0$ for all decision-variables $j$

and

$$\partial L/\partial \lambda_i = 0 \quad \text{for all constraints } i \tag{4.48}$$

are possible local optima.

There is no guarantee that a global optimum solution will be found using calculus-based methods such as this one. Boundary conditions need to be checked. Furthermore, since there is no difference in the Lagrange multipliers procedure for finding a minimum or a maximum solution, one needs to check whether in fact a maximum or minimum is being obtained. In this example, since each net benefit function is concave, a maximum will result.

The meaning of the values of the multipliers $\lambda_i$ at the optimum solution can be derived by manipulation of Equation 4.48. Taking the partial derivative of the Lagrange function, Equation 4.47, with respect to an unknown variable $x_j$ and setting it to zero results in:

$$\partial L/\partial x_j = 0 = \partial F/\partial x_j - \sum_i \lambda_i \partial(g_i(\boldsymbol{X}))/\partial x_j \tag{4.49}$$

Multiplying each term by $\partial x_j$ yields

$$\partial F = \sum_i \lambda_i \partial(g_i(\boldsymbol{X})) \tag{4.50}$$

Dividing each term by $\partial b_k$ associated with a particular constraint, say $k$, defines the meaning of $\lambda_k$.

$$\partial F/\partial b_k = \sum_i \lambda_i \partial(g_i(\boldsymbol{X}))/\partial b_k = \lambda_k \tag{4.51}$$

Equation 4.51 follows from the fact that $\partial(g_i(\boldsymbol{X}))/\partial b_k = 0$ for constraints $i \neq k$ and $\partial(g_i(\boldsymbol{X}))/\partial b_k = 1$ for the constraint $i = k$. The latter is true since $b_i = g_i(\boldsymbol{X})$ and thus $\partial(g_i(\boldsymbol{X})) = \partial b_i$.

Thus from Equation 4.51, each multiplier $\lambda_k$ is the marginal change in the original objective function $F(\boldsymbol{X})$ with respect to a change in the constant $b_k$ associated with the constraint $k$. For non-linear problems it is the slope of the objective function plotted against the value of $b_k$.

Readers can work out a similar proof if a slack or surplus variable, $S_i$, is included in inequality constraints to make them equations. For a less-than-or-equal constraint $g_i(\boldsymbol{X}) \leq b_i$ a squared slack variable $S_i^2$ can be added to the left-hand side to make it an equation $g_i(\boldsymbol{X}) + S_i^2 = b_i$. For a greater-than-or-equal constraint $g_i(\boldsymbol{X}) \geq b_i$ a squared surplus variable $S_i^2$ can be subtracted from the left hand side to make it an equation $g_i(\boldsymbol{X}) - S_i^2 = b_i$. These slack or surplus variables are squared to ensure they are non-negative, and also to make them appear in the differential equations.

$$\partial L/\partial S_i = 0 = -2S_i\lambda_i = S_i\lambda_i \tag{4.52}$$

Equation 4.52 shows that either the slack or surplus variable, $S_i$, or the multiplier, $\lambda_i$, will always be zero. If the value of the slack or surplus variable $S_i$ is non-zero, the constraint is redundant. The optimal solution will not be affected by the constraint. Small changes in the values, $b_i$, of redundant constraints will not change the optimal value of the objective function $F(\mathbf{X})$. Conversely, if the constraint is binding, the value of the slack or surplus variable $S_i$ will be zero. The multiplier $\lambda_i$ can be non-zero if the value of the function $F(\mathbf{X})$ is sensitive to the constraint value $b_i$.

The solution of the set of partial differential Equations 4.52 often involves a trial-and-error process, equating to zero a $\lambda_i$ or a $S_i$ for each inequality constraint and solving the remaining equations, if possible. This tedious procedure, along with the need to check boundary solutions when non-negativity conditions are imposed, detracts from the utility of classical Lagrange multiplier methods for solving all but relatively simple water resources planning problems.

# 4. Dynamic Programming

The water allocation problems in the previous section considered a net-benefit function for each water-using firm. In those examples they were continuous differentiable functions, a convenient attribute if methods based on calculus (such as hill-climbing or Lagrange multipliers) are to be used to find the best solution. In many practical situations these functions may not be so continuous, or so conveniently concave for maximization or convex for minimization, making calculus-based methods for their solution difficult.

A possible solution method for constrained optimization problems containing continuous and/or discontinuous functions of any shape is called *discrete dynamic programming*. Each decision-variable value can assume one of a set of discrete values. For continuous valued objective functions, the solution derived from discrete dynamic programming may therefore be only an approximation of the best one. For all practical purposes this is not a significant limitation, especially if the intervals between the discrete values of the decision-variables are not too large and if simulation modelling is used to refine the solutions identified using dynamic programming.

Dynamic programming is an approach that divides the original optimization problem, with all of its variables, into a set of smaller optimization problems, each of which needs to be solved before the overall optimum solution to the original problem can be identified. The water supply allocation problem, for example, needs to be solved for a range of water supplies available to each firm. Once this is done the particular allocations that maximize the total net benefit can be determined.

## 4.1. Dynamic Programming Networks and Recursive Equations

A network of nodes and links can represent each discrete dynamic programming problem. Dynamic programming methods find the best way to get to any node in that network. The nodes represent possible discrete states that can exist and the links represent the decisions one could make to get from one state to another. Figure 4.6 illustrates a portion of such a network for the three-firm allocation problem shown in Figure 4.1. In this case the total amount of water available, $Q - R$, to all three firms is 10.

Thus, dynamic programming models involve states, stages and decisions. The relationships among states, stages and decisions are represented by networks, such as that shown in Figure 4.6. The states of the system are the nodes and the values of the states are the numbers in the nodes. Each node value in this example is the quantity of water available to allocate to all remaining firms, that is, to all connected links to the right of the node. These state variable values typically represent some existing condition either before making, or after having made, a decision. The stages of the system are the separate columns of linked nodes. The links in this example represent possible allocation decisions for each of the three different firms. Each stage is a separate firm.

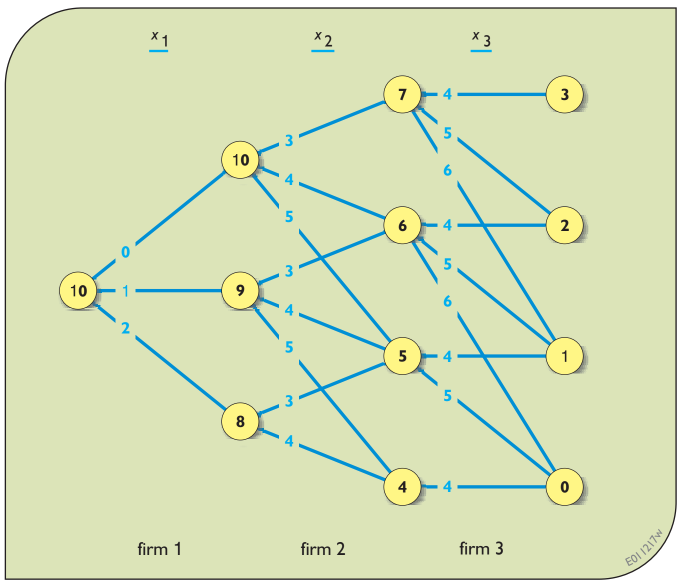



**Figure 4.6.** A network representing some of the possible integer allocations of water to three water-consuming firms $j$. The circles or nodes represent the discrete quantities of water available, and the links represent feasible allocation decisions $x_j$.

Each link connects two nodes, the left node value indicating the state of a system before a decision is made, and the right node value indicating the state of a system after a decision is made. In this case the state of the system is the amount of water available to allocate to the remaining firms.

In the example shown in Figure 4.6, the state and decision-variables are represented by integer values – an admittedly fairly coarse discretization. The total amount of water available, in addition to the amount that must remain in the river, is 10. Note from the first row of Table 4.2 the exact allocation solution is $x_1 = 1.2$, $x_2 = 3.7$, and $x_3 = 5.1$. Normally we wouldn't know this solution before solving for it using dynamic programming, but since we do we can reduce the complexity of the dynamic programming network so that the repetitive process of finding the best solution is clearer. Thus assume the range of $x_1$ is limited to integer values from 0 to 2, the range of $x_2$ is from 3 to 5, and the range of $x_3$ is from 4 to 6. These range limits are imposed here just to reduce the size of the network. In this case, these assumptions will not affect or constrain the optimal solution. If we did not make these assumptions the network would have, after the first column of one node, three columns of 11 nodes, one representing each integer value from 0 to 10. Finer (non-integer) discretizations would involve even more nodes and connecting links.

The links of Figure 4.6 represent the water allocations. Note that the link allocations, the numbers on the links, cannot exceed the amount of water available, that is, the number in the left node. The number in the right node is the quantity of water remaining after an allocation has been made. The value in the right node, state $S_{j+1}$, at the beginning of stage $j+1$, is equal to the value in the left node, $S_j$, less the amount of water, $x_j$, allocated to firm $j$ as indicated on the link. Hence, beginning with a quantity of water $Q - R$ that can be allocated to all three firms, after allocating $x_1$ to Firm 1 what remains is $S_2$:

$$Q - R - x_1 = S_2 \tag{4.53}$$

Allocating $x_2$ to Firm 2, leaves $S_3$.

$$S_2 - x_2 = S_3 \tag{4.54}$$

Finally, allocating $x_3$ to Firm 3 leaves $S_4$.

$$S_3 - x_3 = S_4 \tag{4.55}$$

Figure 4.6 shows the different values of each of these states, $S_j$, and decision-variables $x_j$ beginning with a quantity $Q - R = 10$. Our task is to find the best path through the network, beginning at the left-most node having a state value of 10. To do this we need to know the net benefits we will get associated with all the links (representing the allocation decisions we could make) at each node (state) for each firm (stage).

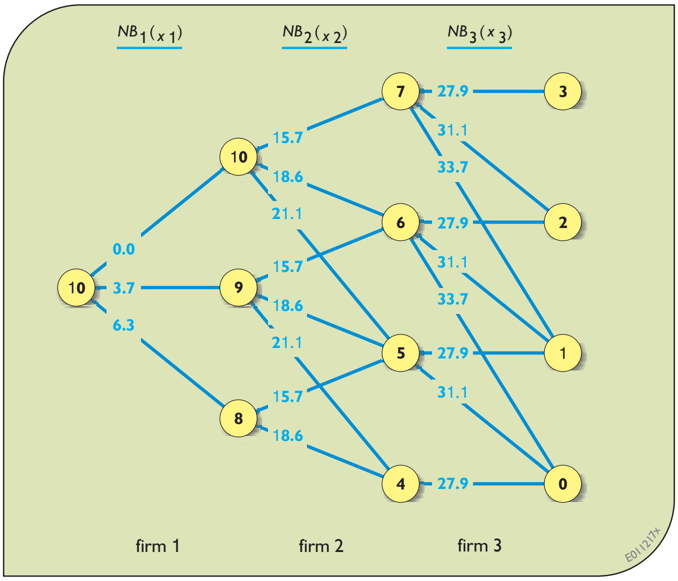



**Figure 4.7.** Network representing integer value allocations of water to three water-consuming firms. The circles or nodes represent the discrete quantities of water available, and the links represent feasible allocation decisions. The numbers on the links indicate the net benefits obtained from these particular allocation decisions.

Figure 4.7 shows the same network as in Figure 4.6; however the numbers on the links represent the net benefits obtained from the associated water allocations. For the three firms $j = 1$, 2 and 3, the net benefits, $NB_j(x_j)$, associated with allocations $x_j$ are:

$$NB_1(x_1) = \text{maximum}(12 - p_1)p_1 - 3(p_1)^{1.30} \quad \text{where } p_1 \leq 0.4(x_1)^{0.9} \tag{4.56}$$

$$NB_2(x_2) = \text{maximum}(20 - 1.5p_2)p_2 - 5(p_2)^{1.20} \quad \text{where } p_2 \leq 0.5(x_2)^{0.8} \tag{4.57}$$

$$NB_3(x_3) = \text{maximum}(28 - 2.5p_3)p_3 - 6(p_3)^{1.15} \quad \text{where } p_3 \leq 0.6(x_3)^{0.7} \tag{4.58}$$

respectively.

The discrete dynamic programming algorithm or procedure is a systematic way to find the best path through this network, or any other suitable network. What makes a network suitable for dynamic programming is the fact that all the nodes can be lined up in a sequence of columns and each link connects a node in one column to another node in the next column of nodes. No link passes over or through any other column(s) of nodes. Links also do not connect nodes in the same column. In addition, the contribution to the overall objective value (in this case, the total net benefits) associated with each discrete decision (link) in any stage or for any firm is strictly a function of the allocation of water to the firm. It is not dependent on the allocation decisions associated with other stages (firms) in the network.

The main challenge in using discrete dynamic programming to solve an optimization problem is to structure the problem so that it fits this dynamic programming network format. Perhaps surprisingly, many water resources planning and management problems do. But it takes practice to become good at converting optimization problems to networks of states, stages and decisions suitable for solution by discrete dynamic programming algorithms.

In this problem the overall objective is to:

$$\text{Maximize} \sum_{j=1}^{3} NB_j(x_j) \tag{4.59}$$

where $NB_j(x_j)$ is the net benefit associated with an allocation of $x_j$ to firm $j$. Equations 4.56, 4.57 and 4.58 define these net benefit functions. As before, the index $j$ represents the particular firm, and each firm is a stage for this problem. Note that the index or subscript used in the objective function often represents an object (like a water-using firm) at a place in space or in a time period. These places or time periods are called the *stages* of a dynamic programming problem. Our task is to find the best path from one stage to the next: in other words, the best allocation decisions for all three firms.

Dynamic programming is called a multi-stage decision-making process. Instead of deciding all three allocations in one single optimization procedure, like Lagrange multipliers, the dynamic programming procedure divides the problem up into many optimization problems, one for each possible discrete state (e.g., amount of water available) in each stage (e.g., for each firm). Given a particular state $S_j$ and stage $j$ – that is, a particular node in the network – what decision (link) $x_j$ will result in the maximum total net benefits, designated as $F_j(S_j)$, given this state $S_j$ for this and all remaining stages or firms $j$, $j+1$, $j+2$ … ? This question must be answered for each node in the network before one can find the overall best set of decisions for each stage: in other words, the best allocations to each firm (represented by the best path through the network) in this example.

Dynamic programming networks can be solved in two ways – beginning at the most right column of nodes or states and moving from right to left, called the *backward-moving* (but forward-looking) algorithm, or beginning at the left most node and moving from left to right, called the *forward-moving* (but backward-looking) algorithm. Both methods will find the best path through the network. In some problems, however, only the backward-moving algorithm produces a useful solution. This is especially relevant when the stages are time periods. We often want to know what we should do next given a particular state we are in, not what we should have just done to get to the particular state we are in. We cannot alter past decisions, but we can, and indeed must, make future decisions. We will revisit this issue when we get to reservoir operation where the stages are time periods.

## 4.2. Backward-Moving Solution Procedure

Consider the network in Figure 4.7. Again, the nodes represent the discrete states – the water available to allocate to all remaining users. The links represent particular discrete allocation decisions. The numbers on the links are the net benefits obtained from those allocations. We want to proceed through the node-link network from the state of 10 at the beginning of the first stage to the end of the network in such a way as to maximize total net benefits. But without looking at all combinations of successive allocations we cannot do this beginning at a state of 10. However, we can find the best solution if we assume we have already made the first two allocations and are at any of the nodes or states at the beginning of the final, third, stage with only one allocation decision remaining. Clearly at each node representing the water available to allocate to the third firm, the best decision is to pick the allocation (link) having the largest net benefits.

Denoting $F_3(S_3)$ as the maximum net benefits we can achieve from the remaining amount of water $S_3$, then for each discrete value of $S_3$ we can find the $x_3$ that maximizes $F_3(S_3)$. Those shown in Figure 4.7 include:

$$F_3(7) = \text{Maximum}\{NB_3(x_3)\}$$
$$x_3 \leq 7, \text{ the total flow available.}$$

$$\begin{aligned} &4 \leq x_3 \leq 6, \text{ the allowable range of allocation decisions.} \\ &= \text{Maximum}\{27.9, 31.1, 33.7\} = 33.7 \\ &\text{when } x_3 = 6 \end{aligned} \tag{4.60}$$

$$\begin{aligned} F_3(6) &= \text{Maximum}\{NB_3(x_3)\} \\ &\quad x_3 \leq 6 \\ &\quad 4 \leq x_3 \leq 6 \\ &= \text{Maximum}\{27.9, 31.1, 33.7\} = 33.7 \\ &\quad \text{when } x_3 = 6 \end{aligned} \tag{4.61}$$

$$\begin{aligned} F_3(5) &= \text{Maximum}\{NB_3(x_3)\} \\ &\quad x_3 \leq 5 \\ &\quad 4 \leq x_3 \leq 6 \\ &= \text{Maximum}\{27.9, 31.1\} = 31.1 \\ &\quad \text{when } x_3 = 5 \end{aligned} \tag{4.62}$$

$$\begin{aligned} F_3(4) &= \text{Maximum}\{NB_3(x_3)\} \\ &\quad x_3 \leq 4 \\ &\quad 4 \leq x_3 \leq 6 \\ &= \text{Maximum}\{27.9\} = 27.9 \quad \text{when } x_3 = 4 \end{aligned} \tag{4.63}$$

These computations are shown on the network in Figure 4.8. Note that there are no benefits to be obtained after the third allocation, so the decision to be made for each node or state prior to allocating water to Firm 3 is simply that which maximizes the net benefits derived from that last (third) allocation. In Figure 4.8 the links representing the decisions or allocations that result in the largest net benefits are shown with arrows.

Having computed the maximum net benefits, $F_3(S_3)$, associated with each initial state $S_3$ for Stage 3, we can now move backward (to the left) to the discrete states $S_2$ at the beginning of the second stage. Again, these states represent the quantity of water available to allocate to Firms 2 and 3. Denote $F_2(S_2)$ as the maximum total net benefits obtained from the two remaining allocations $x_2$ and $x_3$ given the quantity $S_2$ water available. The best $x_2$ depends not only on the net benefits obtained from the allocation $x_2$ but also on the maximum net benefits obtainable after that, namely the just-calculated $F_3(S_3)$ associated with the state $S_3$ that results from the initial state $S_2$ and a decision $x_2$. As defined in Equation 4.54, this final state $S_3$ in Stage 2 obviously equals $S_2 - x_2$. Hence for those nodes at the beginning of Stage 2 shown in Figure 4.8:

$$\begin{aligned} F_2(10) &= \text{Maximum}\{NB_2(x_2) + F_3(S_3 = 10 - x_2)\} \\ &\quad x_2 \leq 10 \\ &\quad 3 \leq x_2 \leq 5 \\ &= \text{Maximum}\{15.7 + 33.7, 18.6 + 33.7, \\ &\quad 21.1 + 31.1\} = 52.3 \quad \text{when } x_2 = 4 \end{aligned} \tag{4.64}$$

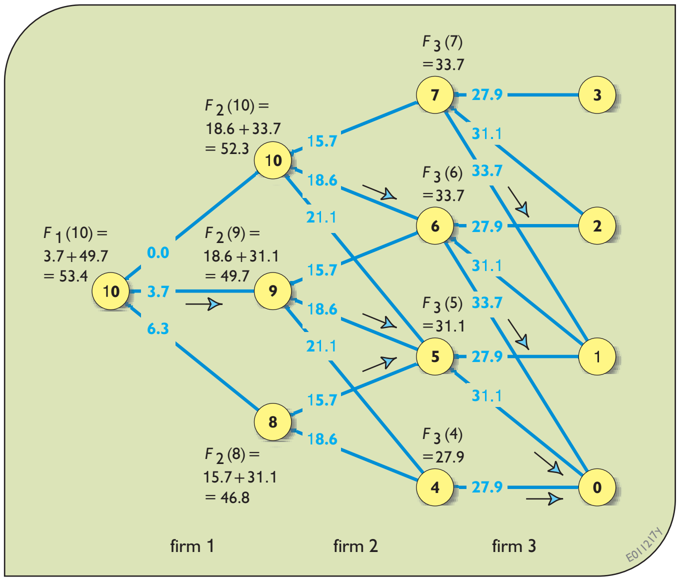



**Figure 4.8.** Using the backward-moving dynamic programming method for finding the maximum remaining net benefits, $F_j(S_j)$, and optimal allocations (denoted by the arrows on the links) for each state in Stage 3, then for each state in Stage 2 and finally for the initial state in Stage 1 to obtain the optimum allocation policy and maximum total net benefits, $F_1(10)$. The minimum flow to remain in the river, $R$, is in addition to the ten units available for allocation and is not shown in this network.

$$\begin{aligned} F_2(9) &= \text{Maximum}\{NB_2(x_2) + F_3(S_3 = 9 - x_2)\} \\ &\quad x_2 \leq 9 \\ &\quad 3 \leq x_2 \leq 5 \\ &= \text{Maximum}\{15.7 + 33.7, 18.6 + 31.1, \\ &\quad 21.1 + 27.9\} = 49.7 \quad \text{when } x_2 = 4 \end{aligned} \tag{4.65}$$

$$\begin{aligned} F_2(8) &= \text{Maximum}\{NB_2(x_2) + F_3(S_3 = 8 - x_2)\} \\ &\quad x_2 \leq 8 \\ &\quad 3 \leq x_2 \leq 5 \\ &= \text{Maximum}\{15.7 + 31.1, 18.6 + 27.9\} = 46.8 \\ &\quad \text{when } x_2 = 3 \end{aligned} \tag{4.66}$$

These maximum net benefit functions, $F_2(S_2)$, could be calculated for the remaining discrete states from 7 to 0.

Having computed the maximum net benefits obtainable for each discrete state at the beginning of Stage 2, that is, all the $F_2(S_2)$ values, we can move backward or left to the beginning of Stage 1. For this problem there is only one state, the state of 10 we are actually in before making any allocations to any of the firms. In this case, the maximum net benefits, $F_1(10)$, we can obtain from all three allocations given 10 units of water available, is

$$F_1(10) = \text{Maximum}\{NB_1(x_1) + F_2(S_2 = 10 - x_1)\} \quad (4.67)$$
$$x_1 \leq 10$$
$$0 \leq x_1 \leq 2$$
$$= \text{Maximum}\{0 + 52.3, 3.7 + 49.7, 6.3 + 46.8\} = 53.4 \quad \text{when } x_1 = 1$$

Equation 4.67 is another way of expressing Equation 4.59. Their values are the same. It is the maximum net benefits obtainable from allocating the available 10 units of water. From Equation 4.67 we know that we will get a maximum of 53.4 net benefits if we allocate 1 unit of water to Firm 1. This leaves 9 units of water to allocate to the two remaining firms. This is our optimal state at the beginning of Stage 2. Given a state of 9 at the beginning of Stage 2, we see from Equation 4.65 that we should allocate 4 units of water to Firm 2. This leaves 5 units of water for Firm 3. Given a state of 5 at the beginning of Stage 3, Equation 4.62 tells us we should allocate all 5 units to Firm 3. All this is illustrated in Figure 4.8.

Compare this discrete solution with the continuous one defined by Lagrange multipliers as shown in Table 4.2. The exact solution, to the nearest tenth, is 1.2, 3.7, and 5.1 for $x_1$, $x_2$ and $x_3$ respectively. The solution just derived from discrete dynamic programming that assumed only integer values is 1, 4, and 5 respectively.

To summarize, a dynamic programming model was developed for the following problem:

$$\text{Maximize Net\_benefit} \quad (4.68)$$

Subject to:

$$\text{Net\_benefit} = \text{Total\_return} - \text{Total\_cost} \quad (4.69)$$

$$\text{Total\_return} = (12 - p_1)p_1 + (20 - 1.5p_2)p_2 + (28 - 2.5p_3)p_3 \quad (4.70)$$

$$\text{Total\_cost} = 3(p_1)^{1.30} + 5(p_2)^{1.20} + 6(p_3)^{1.15} \quad (4.71)$$

$$p_1 \leq 0.4(x_1)^{0.9} \quad (4.72)$$

$$p_2 \leq 0.5(x_2)^{0.8} \quad (4.73)$$

$$p_3 \leq 0.6(x_3)^{0.7} \quad (4.74)$$

$$x_1 + x_2 + x_3 \leq 10 \quad (4.75)$$

The discrete dynamic programming version of this problem required discrete states $S_j$ representing the amount of water available to allocate to firms $j, j+1, \ldots$, . It required discrete allocations $x_j$. Next it required the calculation of the maximum net benefits, $F_j(S_j)$, that could be obtained from all firms $j$, beginning with Firm 3, and proceeding backwards as indicated in Equations 4.76 to 4.78.

$$F_3(S_3) = \text{maximum}\{NB_3(x_3)\} \text{ over all } x_3 \leq S_3, \quad \text{for all discrete } S_3 \text{ values between 0 and 10} \quad (4.76)$$

$$F_2(S_2) = \text{maximum}\{NB_2(x_2) + F_3(S_3)\} \text{ over all } x_2 \leq S_2 \text{ and } S_3 = S_2 - x_2, \quad 0 \leq S_2 \leq 10 \quad (4.77)$$

$$F_1(S_1) = \text{maximum}\{NB_1(x_1) + F_2(S_2)\} \text{ over all } x_1 \leq S_1 \text{ and } S_2 = S_1 - x_1 \text{ and } S_1 = 10 \quad (4.78)$$

To solve for $F_1(S_1)$ and each optimal allocation $x_j$ we must first solve for all values of $F_3(S_3)$. Once these are known we can solve for all values of $F_2(S_2)$. Given these $F_2(S_2)$ values, we can solve for $F_1(S_1)$. Equations 4.76 need to be solved before Equations 4.77 can be solved, and Equations 4.77 need to be solved before Equations 4.78 can be solved. They need not be solved simultaneously, and they cannot be solved in reverse order. These three equations are called recursive equations. They are defined for the backward-moving dynamic programming solution procedure.

There is a correspondence between the non-linear optimization model defined by Equations 4.68 to 4.75 and the dynamic programming model defined by the recursive Equations 4.76, 4.77 and 4.78. Note that $F_3(S_3)$ in Equation 4.76 is the same as:

$$F_3(S_3) = \text{Maximum } NB_3(x_3) \quad (4.79)$$

Subject to:

$$x_3 \leq S_3 \quad (4.80)$$

where $NB_3(x_3)$ is defined in Equation 4.58.

Similarly, $F_2(S_2)$ in Equation 4.62 is the same as:

$$F_2(S_2) = \text{Maximum } NB_2(x_2) + NB_3(x_3) \quad (4.81)$$

Subject to:

$$x_2 + x_3 \leq S_2 \quad (4.82)$$

where $NB_2(x_2)$ and $NB_3(x_3)$ are defined in Equations 4.57 and 4.58.

Finally, $F_1(S_1)$ in Equation 4.63 is the same as:

$$F_1(S_1) = \text{Maximum } NB_1(x_1) + NB_2(x_2) + NB_3(x_3) \quad (4.83)$$

Subject to:

$$x_1 + x_2 + x_3 \leq S_1 = 10 \quad (4.84)$$

where $NB_1(x_1)$, $NB_2(x_2)$ and $NB_3(x_3)$ are defined in Equations 4.56, 4.57 and 4.58.

Alternatively, $F_3(S_3)$ in Equation 4.76 is the same as:

$$F_3(S_3) = \text{Maximum}(28 - 2.5p_3)p_3 - 6(p_3)^{1.15} \quad (4.85)$$

Subject to:

$$p_3 \leq 0.6(x_3)^{0.7} \quad (4.86)$$

$$x_3 \leq S_3 \quad (4.87)$$

Similarly, $F_2(S_2)$ in Equation 4.77 is the same as:

$$F_2(S_2) = \text{Maximum}(20 - 1.5p_2)p_2 + (28 - 2.5p_3) \times p_3 - 5(p_2)^{1.20} - 6(p_3)^{1.15} \quad (4.88)$$

Subject to:

$$p_2 \leq 0.5(x_2)^{0.8} \quad (4.89)$$

$$p_3 \leq 0.6(x_3)^{0.7} \quad (4.90)$$

$$x_2 + x_3 \leq S_2 \quad (4.91)$$

Finally, $F_1(S_1)$ in Equation 4.78 is the same as:

$$F_1(S_1) = \text{Maximum}(12 - p_1)p_1 + (20 - 1.5p_2)p_2 + (28 - 2.5p_3)p_3 - [3(p_1)^{1.30} + 5(p_2)^{1.20} + 6(p_3)^{1.15}] \quad (4.92)$$

Subject to:

$$p_1 \leq 0.4(x_1)^{0.9} \quad (4.93)$$

$$p_2 \leq 0.5(x_2)^{0.8} \quad (4.94)$$

$$p_3 \leq 0.6(x_3)^{0.7} \quad (4.95)$$

$$x_1 + x_2 + x_3 \leq S_1 = 10 \quad (4.96)$$

The transition function of dynamic programming defines the relationship between two successive states $S_j$ and $S_{j+1}$ and the decision $x_j$. In the above example these transition functions are defined by Equations 4.53, 4.54 and 4.55, or, in general terms for all firms $j$, by:

$$S_{j+1} = S_j - x_j \quad (4.97)$$

## 4.3. Forward-Moving Solution Procedure

We have just described the backward-moving dynamic programming algorithm. In that approach at each node (state) in each stage we calculated the best value of the objective function that can be obtained from all further or remaining decisions. Alternatively one can proceed forward, that is, from left to right, through a dynamic programming network. For the forward-moving algorithm at each node we need to calculate the best value of the objective function that could be obtained from all past decisions leading to that node or state. In other words, we need to find how best to get to each state $S_{j+1}$ at the end of each stage $j$.

Returning to the allocation example, define $f_j(S_{j+1})$ as the maximum net benefits from the allocation of water to firms 1, 2, … , $j$, given a state $S_{j+1}$ after having made those allocations. For this example we begin the forward-moving, but backward-looking, process by selecting each of the ending states in the first stage $j = 1$ and finding the best way to have arrived at (or to have achieved) those ending states. Since in this example there is only one way to get to each of those states, as shown in Figure 4.6 or 4.7, the allocation decisions are obvious.

$$f_1(S_2) = \text{maximum}\{NB_1(x_1)\}$$
$$x_1 = 10 - S_2 \quad (4.98)$$

Hence, $f_1(S_2)$ is simply $NB_1(10 - S_2)$. Once the values for all $f_1(S_2)$ are known for all discrete $S_2$ between 0 and 10, move forward (to the right) to the end of Stage 2 and find the best allocations to have made given each final state $S_3$.

$$f_2(S_3) = \text{maximum}\{NB_2(x_2) + f_1(S_2)\}$$
$$0 \leq x_2 \leq 10 - S_3$$
$$S_2 = S_3 + x_2 \quad (4.99)$$

Once the values of all $f_2(S_3)$ are known for all discrete states $S_3$ between 0 and 10, move forward to Stage 3 and find the best allocations to have made given each final state $S_4$.

$$f_3(S_4) = \text{maximum}\{NB_3(x_3) + f_2(S_3)\}$$

for all discrete $S_4$ between 0 and 10.

$$0 \leq x_3 \leq 10 - S_4$$
$$S_3 = S_4 + x_3 \quad (4.100)$$

Figure 4.9 illustrates a portion of the network represented by Equations 4.98 through 4.100, and the $f_j(S_{j+1})$ values.

From Figure 4.9, note the highest total net benefits are obtained by ending with 0 remaining water at the end of Stage 3. The arrow tells us that if we are to get to that state optimally, we should allocate 5 units of water to Firm 3. Thus we must begin Stage 3, or end Stage 2, with $10 - 5 = 5$ units of water. To get to this state at the end of Stage 2 we should allocate 4 units of water to Firm 2.

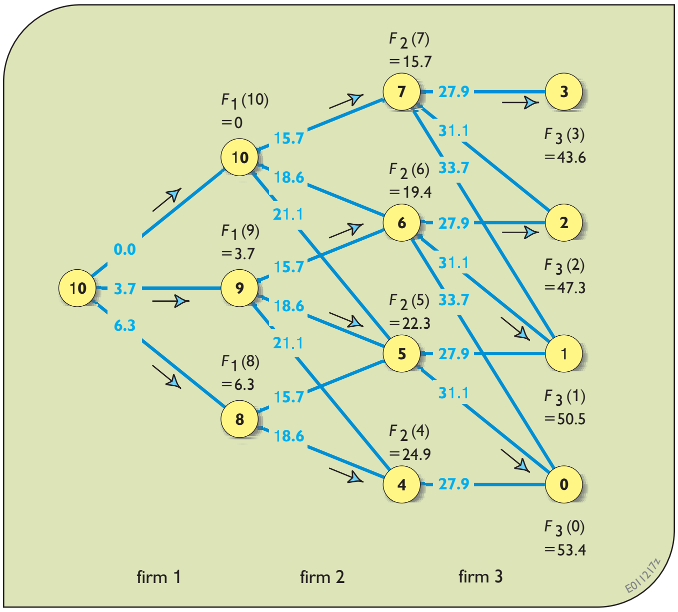



**Figure 4.9.** Using the forward-moving dynamic programming method for finding the maximum accumulated net benefits, $F_j(S_{j+1})$, and optimal allocations (denoted by the arrows on the links) that should have been made to reach each ending state, beginning with the ending states in Stage 1, then for each ending state in Stage 2 and finally for the ending states in Stage 3.

The arrow also tells us we should have had 9 units of water available at the end of Stage 1. Given this state of 9 at the end of Stage 1, the arrow tells us we should allocate 1 unit of water to Firm 1. This is the same allocation policy as obtained using the backward-moving algorithm.

## 4.4. Numerical Solutions

The application of discrete dynamic programming to most practical problems will require writing some software. There are no general dynamic programming computer programs available that will solve all dynamic programming problems. Thus any user of dynamic programming will need to write a computer program to solve a particular problem. Most computer programs written for solving specific dynamic programming problems create and store the solutions of the recursive equations in tables. Each stage is a separate table, as shown in Tables 4.3, 4.4 and 4.5 for this example water allocation problem. These tables apply to only a part of the entire problem, namely that part of the network shown in Figures 4.8 and 4.9. The backward solution procedure is used.

Table 4.3 contains the solutions of Equations 4.60 to 4.63 for the third stage. Table 4.4 contains the solutions of Equations 4.64 to 4.66 for the second stage. Table 4.5 contains the solution of Equation 4.67 for the first stage.

From Table 4.5 we see that, given 10 units of water available, we will obtain 53.4 net benefits and to get this we should allocate 1 unit to Firm 1. This leaves 9 units of water for the remaining two allocations. From Table 4.4 we see that for a state of 9 units of water available we should allocate 4 units to Firm 2. This leaves 5 units. From Table 4.3 for a state of 5 units of water available we see we should allocate all 5 of them to Firm 3.

Performing these calculations for various discrete total amounts of water available, say from 0 to 38 in this example, will define an allocation policy (such as the one shown in Figure 4.6 for a different allocation problem) for situations when the total amount of water is less than that desired by all the firms. This policy can then be simulated using alternative time series of available amounts of water, such as streamflows, to obtain estimates of the time series (or statistical measures of those time series) of net benefits obtained by each firm, assuming the allocation policy is followed over time.

**Table 4.3.** Computing the values of $F_3(S_3)$ and optimal allocations $x_3$ for all states $S_3$ in Stage 3.

remaining net benefits $NB_3(S_3)$

| | decisions: $x_3$ | | | | |
|---|---|---|---|---|---|
| state $S_3$ | 4 | 5 | 6 | $F_3(S_3)$ | $x_3$ |
| 7 | 27.9 | 31.1 | 33.7 | 33.7 | 6 |
| 6 | 27.9 | 31.1 | 33.7 | 33.7 | 6 |
| 5 | 27.9 | 31.1 | --- | 31.1 | 5 |
| 4 | 27.9 | --- | --- | 27.9 | 4 |

E020820j

Table 4.4. Computing the values of $F_2(S_2)$ and optimal allocations $x_2$ for all states $S_2$ in Stage 2.

remaining net benefits $NB_2(S_2) + F_3(S_3 = S_2 - x_2)$

| state $S_2$ | decisions: $x_2$ 3 | 4 | 5 | $F_2(S_2)$ | $x_2$ |
|---|---|---|---|---|---|
| 10 | 15.7 + 33.7 | 18.6 + 33.7 | 21.1 + 31.1 | 52.3 | 4 |
| 9 | 15.7 + 33.7 | 18.6 + 31.1 | 21.1 + 27.9 | 49.7 | 4 |
| 8 | 15.7 + 31.1 | 18.6 + 27.9 | --- | 46.8 | 3 |

E020820k

Table 4.5. Computing the values of $F_1(S_1)$ and optimal allocations $x_1$ for all states $S_1$ in Stage 1.

remaining net benefits $NB_1(S_1) + F_2(S_2 = S_1 - x_1)$

| state $S_2$ | decisions: $x_2$ 0 | 1 | 2 | $F_2(S_2)$ | $x_2$ |
|---|---|---|---|---|---|
| 10 | 0 + 52.3 | 3.7 + 49.7 | 6.3 + 46.8 | 53.4 | 1 |

E020820m

## 4.5. Dimensionality

One of the limitations of dynamic programming is handling multiple state variables. In our water allocation example we had only one state variable: the total amount of water available. We could have enlarged this problem to include other types of resources the firms require to make their products. Each of these state variables would need to be discretized. If, for example, only $m$ discrete values of each state variable are considered, for $n$ different state variables (e.g., types of resources) there are $m^n$ different combinations of state variable values to consider at each stage. As the number of state variables increases, the number of discrete combinations of state variable values increases exponentially. This is called dynamic programming's '*curse of dimensionality*'. It has motivated many researchers to search for ways of reducing the number of possible discrete states required to find an optimal solution to large multi-state problems.

## 4.6. Principle of Optimality

The solution of dynamic programming models or networks is based on a principal of optimality (Bellman, 1957). The backward-moving solution algorithm is based on the principal that no matter what the state and stage (i.e., the particular node you are at), an optimal policy is one that proceeds forward from that node or state and stage optimally. The forward-moving solution algorithm is based on the principal that no matter what the state and stage (i.e., the particular node you are at), an optimal policy is one that has arrived at that node or state and stage in an optimal manner.

This 'principle of optimality' is a very simple concept but requires the formulation of a set of recursive equations at each stage. It also requires that either in the last stage ($j = J$) for a backward-moving algorithm, or in the first stage ($j = 1$) for a forward-moving algorithm, the future value functions, $F_{J+1}(S_{J+1})$, associated with the ending state variable values, or past value functions, $f_0(S_1)$, associated with the beginning state variable values, respectively, all equal some known value. Usually that value is 0 but not always. This condition is required in order to begin the process of solving each successive recursive equation.

## 4.7. Additional Applications

Among the common dynamic programming applications in water resources planning are water allocations to multiple uses, infrastructure capacity expansion, and reservoir operation. The previous three-user water allocation problem (Figure 4.1) illustrates the first type of application. The other two applications are presented below.

### 4.7.1. Capacity Expansion

How much infrastructure should be built, when and why? Consider a municipality that must plan for the future expansion of its water supply system or some component of that system, such as a reservoir, aqueduct, or treatment plant. The capacity needed at the end of each future period $t$ has been estimated to be $D_t$. The cost, $C_t(s_t, x_t)$, of adding capacity $x_t$ in each period $t$ is a function of that added capacity as well as of the existing capacity $s_t$ at the beginning of the period. The planning problem is to find that time sequence of capacity expansions that minimizes the present value of total future costs while meeting the predicted capacity demand requirements. This is the usual capacity-expansion problem.

This problem can be written as an optimization model: The objective is to minimize the present value of the total cost of capacity expansion.

$$\text{Minimize } \sum_t C_t(s_t, x_t) \tag{4.101}$$

where $C_t(s_t, x_t)$ is the present value of the cost of capacity expansion $x_t$ in period $t$ given an initial capacity of $s_t$.

The constraints of this model define the minimum required final capacity in each period $t$, or equivalently the next period's initial capacity, $s_{t+1}$, as a function of the known existing capacity $s_1$ and each expansion $x_t$ up through period $t$.

$$s_{t+1} = s_t + \sum_{\tau=1}^{t} x_\tau \quad \text{for } t = 1, 2, \dots, T \tag{4.102}$$

Alternatively these equations may be expressed by a series of continuity relationships:

$$s_{t+1} = s_t + x_t \quad \text{for } t = 1, 2, \dots, T \tag{4.103}$$

In this problem, the constraints must also ensure that the actual capacity $s_{t+1}$ at the end of each future period $t$ is no less than the capacity required $D_t$ at the end of that period.

$$s_{t+1} \geq D_t \quad \text{for } t = 1, 2, \dots, T \tag{4.104}$$

There may also be constraints on the possible expansions in each period defined by a set $\boldsymbol{\Omega}_t$ of feasible capacity additions in each period $t$:

$$x_t \in \boldsymbol{\Omega}_t \tag{4.105}$$

Figure 4.10 illustrates this type of capacity-expansion problem. The question is how much capacity to add and when. It is a significant problem for several reasons. One is that the cost functions $C_t(s_t, x_t)$ typically exhibit

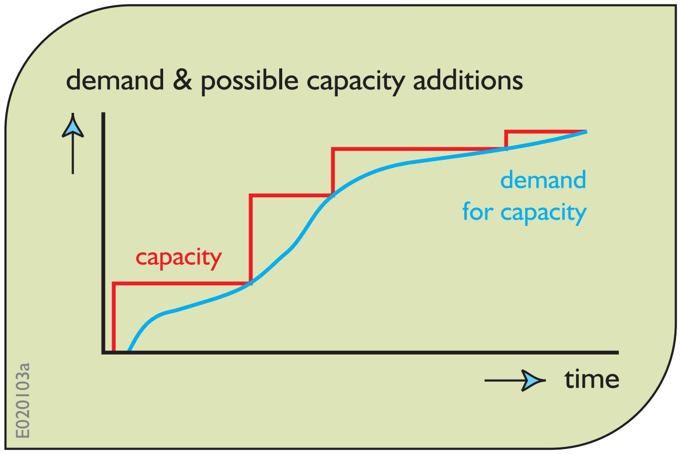



**Figure 4.10.** A demand projection (solid line) and a possible capacity-expansion schedule for meeting that projected demand over time.

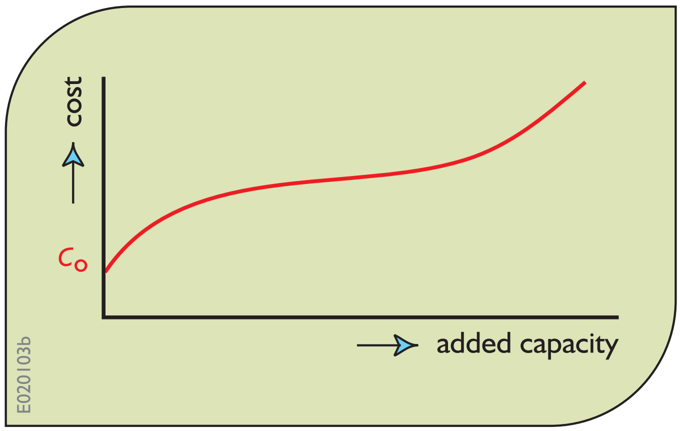



**Figure 4.11.** Typical cost function for additional capacity given an existing capacity. The cost function shows the fixed costs, $C_0$, required if additional capacity is to be added, and the economies of scale associated with the concave portion of the cost function.

fixed costs and economies of scale, as illustrated in Figure 4.11. Each time any capacity is added there are fixed as well as variable costs incurred. Fixed and variable costs that show economies of scale (decreasing average costs associated with increasing capacity additions) motivate the addition of excess capacity, capacity not needed immediately but expected to be needed eventually to meet an increased demand for additional capacity in the future.

The problem is also important because any estimates made today of future demands, costs and interest rates are likely to be wrong. The future is uncertain. Its uncertainties increase the further we look into the future. Capacity-expansion planners need to consider the future if their plans are to be cost-effective. Just how far into the future do they need to look? And what about the uncertainty in all future estimates? These questions will be addressed after showing how the problem can be solved for any fixed-planning horizon and estimates of future demands, interest rates and costs.

The constrained optimization model defined by Equations 4.101 to 4.105 can be restructured as a multi-stage decision-making process and solved using either a forward or backward-moving discrete dynamic programming solution procedure. The stages of the model will be the time periods $t$. The states will be either the capacity $s_{t+1}$ at the end of a stage or period $t$ if a forward-moving solution procedure is adopted, or the capacity $s_t$, at the beginning of a stage or period $t$ if a backward-moving solution procedure is used.

A network of possible discrete capacity states and decisions can be superimposed onto the demand projection of Figure 4.10, as shown in Figure 4.12.

The solid blue circles in Figure 4.12 represent possible discrete states, $S_t$, of the system, the amounts of additional capacity existing at the end of each period $t-1$ or equivalently at the beginning of period $t$.

Consider first a forward-moving dynamic programming algorithm. To implement this, define $f_t(s_{t+1})$ as the minimum cost of achieving a capacity $s_{t+1}$, at the end of period $t$. Since at the beginning of the first period $t = 1$, the accumulated least cost is 0, $f_0(s_1) = 0$.

Hence, for each final discrete state $s_2$ in stage $t = 1$ ranging from $D_1$ to the maximum demand $D_T$, define

$$f_1(s_2) = \min\{C_1(s_1, x_1)\} \text{ in which the discrete } x_1 = s_2 \text{ and } s_1 = 0 \tag{4.106}$$

Moving to stage $t = 2$, for the final discrete states $s_3$ ranging from $D_2$ to $D_T$,

$$f_2(s_3) = \min\{C_2(s_2, x_2) + f_1(s_2)\} \text{ over all discrete } x_2 \text{ between 0 and } s_3 - D_1$$

$$\text{where } s_2 = s_3 - x_2 \tag{4.107}$$

Moving to stage $t = 3$, for the final discrete states $s_4$ ranging from $D_3$ to $D_T$,

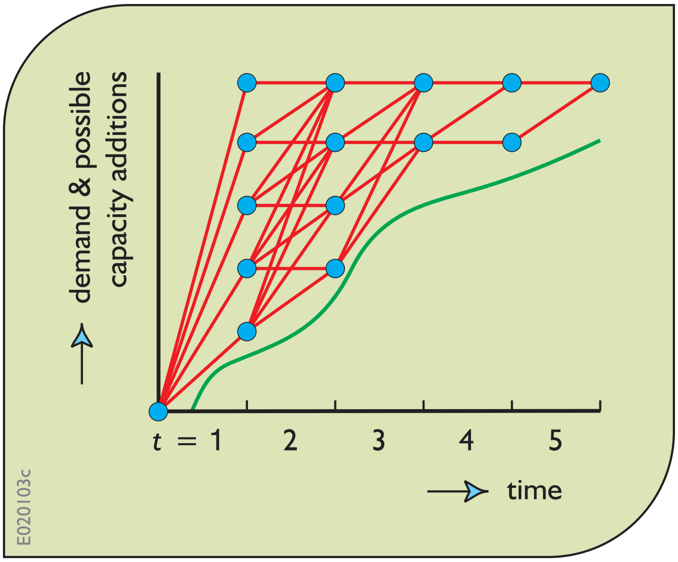



**Figure 4.12.** Network of discrete capacity-expansion decisions (links) that meet the projected demand.

$$f_3(s_4) = \min\{C_3(s_3, x_3) + f_2(s_3)\} \text{ over all discrete } x_3 \text{ between 0 and } s_4 - D_2$$

$$\text{where } s_3 = s_4 - x_3 \tag{4.108}$$

In general for all stages $t$ between the first and last:

$$f_t(s_{t+1}) = \min\{C_t(s_t, x_t) + f_{t-1}(s_t)\} \text{ over all discrete } x_t \text{ between 0 and } s_{t+1} - D_{t-1}$$

$$\text{where } s_t = s_{t+1} - x_t \tag{4.109}$$

For the last stage $t = T$ and for the final discrete state $s_{T+1} = D_T$,

$$f_T(s_{T+1}) = \min\{C_T(s_T, x_T) + f_{T-1}(s_T)\} \text{ over all discrete } x_T \text{ between 0 and } D_T - D_{T-1}$$

$$\text{where } s_T = s_{T+1} - x_T \tag{4.110}$$

The value of $f_T(s_{T+1})$ is the minimum present value of the total cost of meeting the demand for $T$ time periods. To identify the sequence of capacity-expansion decisions that results in this minimum present value of the total cost requires backtracking to collect the set of best decisions $x_t$ for all stages $t$. A numerical example will illustrate this.

## A numerical example

Consider the five-period capacity-expansion problem shown in Figure 4.12. Figure 4.13 is the same network

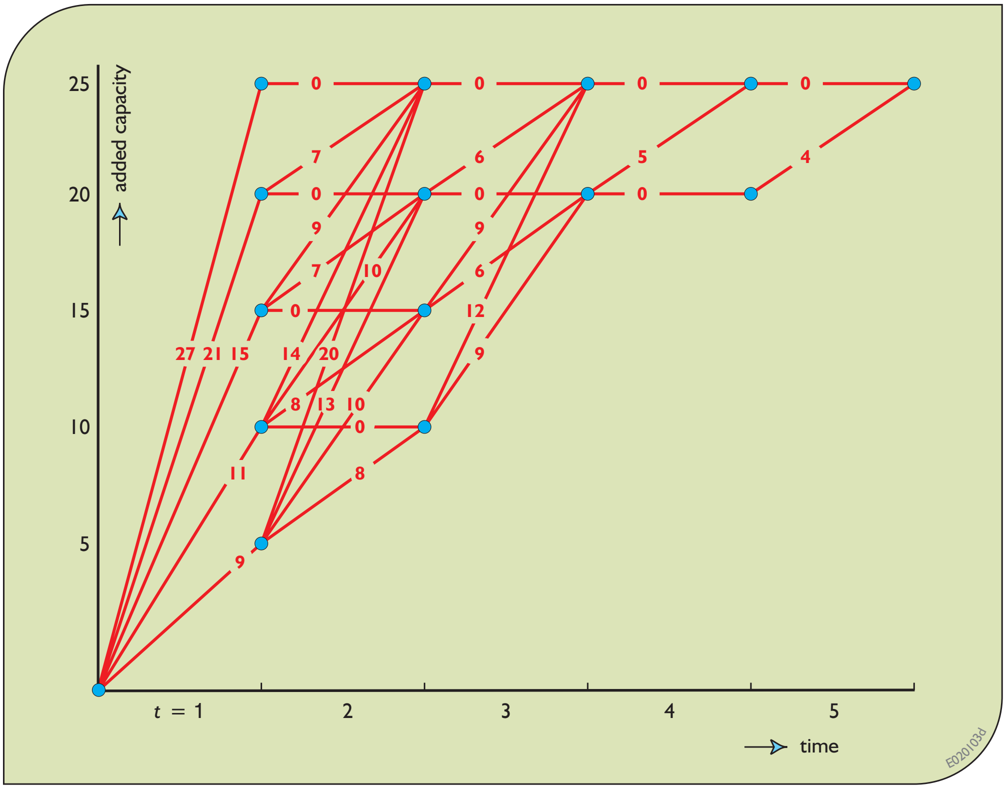



**Figure 4.13.** A discrete capacity-expansion network showing the present value of the expansion costs associated with each feasible expansion decision. Finding the best path through the network can be done using forward or backward-moving discrete dynamic programming.

with the present value of the expansion costs on each link. The values of the states, the existing capacities, represented by the nodes, are shown on the left vertical axis. The capacity-expansion problem is solved on Figure 4.14 using the forward-moving algorithm.

From the forward-moving solution to the dynamic programming problem shown in Figure 4.14, the present value of the cost of the optimal capacity-expansion schedule is 23 units of money. Backtracking (moving left against the arrows) from the farthest right node, this schedule adds 10 units of capacity in period $t = 1$, and 15 units of capacity in period $t = 3$.

Next consider the backward-moving algorithm applied to this capacity-expansion problem. The general recursive equation for a backward-moving solution is

$$F_t(s_t) = \text{minimum}\{C_t(s_t, x_t) + F_{t+1}(s_{t+1})\} \text{ over all discrete } x_t \text{ from } D_t - s_t \text{ to } D_T - s_t$$

$$\text{for all discrete states } s_t \text{ from } D_{t-1} \text{ to } D_T \qquad (4.111)$$

where $F_{T+1}(D_T) = 0$ and as before each cost function is the discounted cost.

Once again, as shown in Figure 4.15, the minimum total present value cost is 23 if 10 units of additional capacity are added in period $t = 1$ and 15 in period $t = 3$.

Now we look to the question of the uncertainty of future demands, $D_t$, discounted costs, $C_t(s_t, x_t)$, as well as to the fact that the planning horizon $T$ is only 5 time periods. Of importance is just how these uncertainties and finite planning horizon affect our decisions. While the model gives us a time series of future capacity-expansion decisions for the next 5 time periods, what is important to decision-makers is what additional capacity to add now, not what capacity to add in future periods. Does the uncertainty of future demands and costs and the 5-period planning horizon affect this first decision, $x_1$? This is the question to address. If the answer is no, then one can place some confidence in the value of $x_1$. If yes, then more study may be warranted to determine which demand and cost scenario to assume, or, if applicable, how far into the future to extend the planning horizon.

Future capacity-expansion decisions in Periods 2, 3 and so on can be based on updated information and analyses carried out closer to the time those decisions are to be made. At those times, the forecast demands and

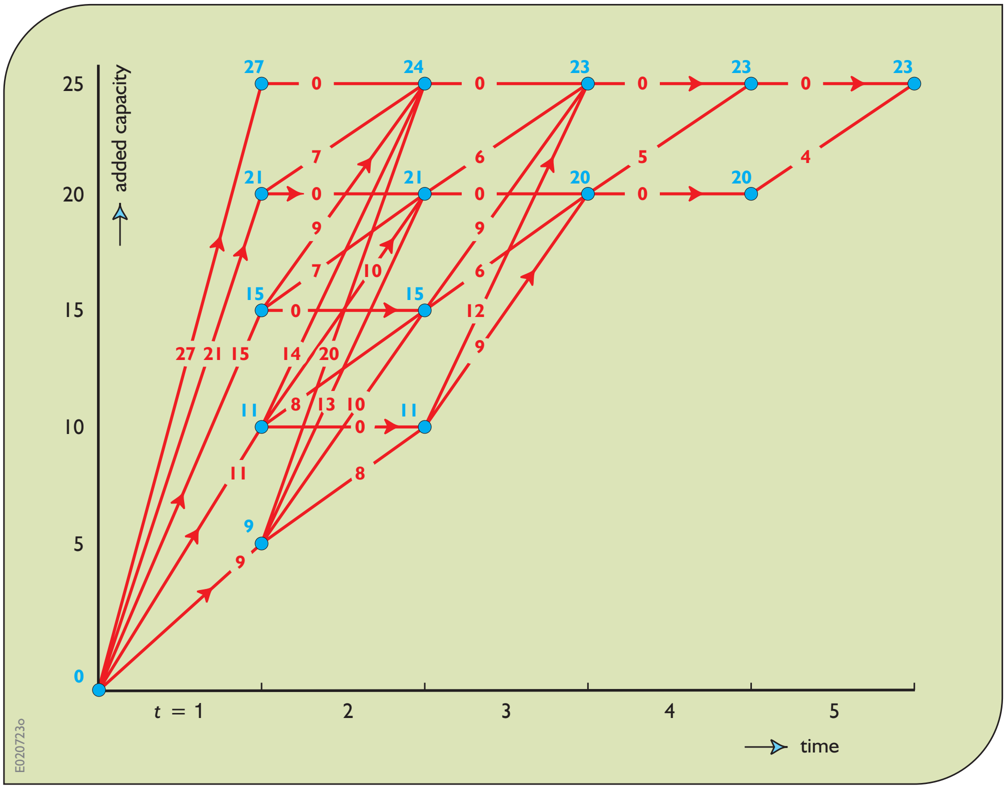



**Figure 4.14.** A capacity-expansion example, showing the results of a forward-moving dynamic programming algorithm. The numbers next to the nodes are the minimum cost to have reached that particular state at the end of the particular time period *t*.

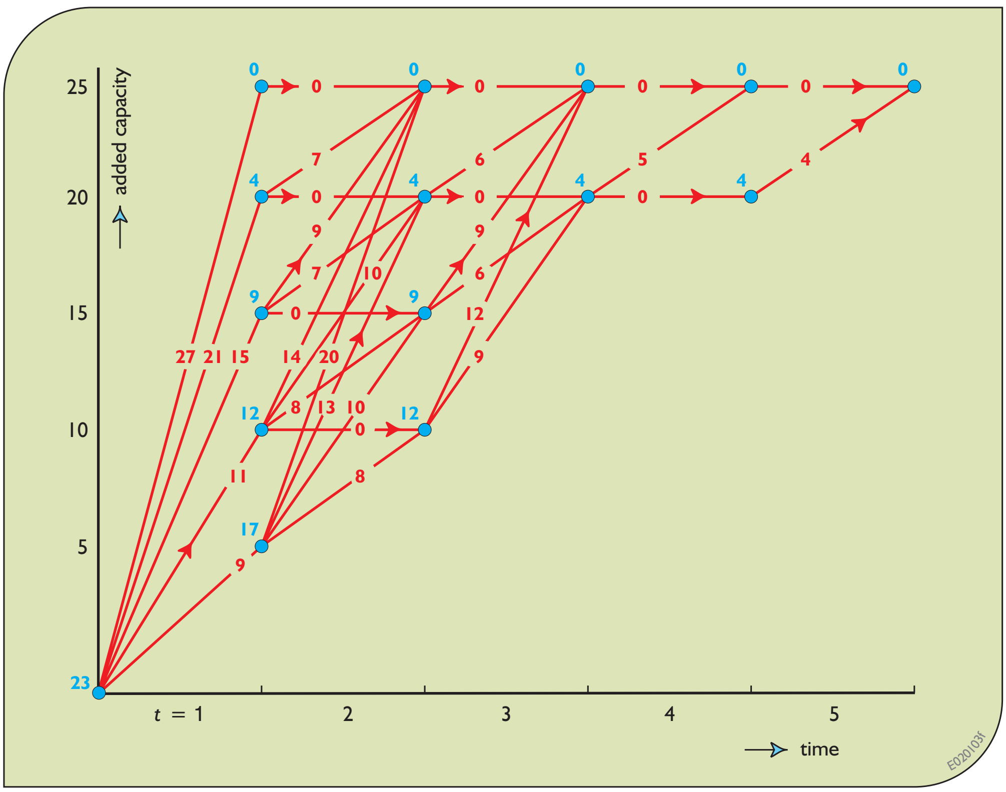



**Figure 4.15.** A capacity-expansion example, showing the results of a backward-moving dynamic programming algorithm. The numbers next to the nodes are the minimum remaining cost to have the particular capacity required at the end of the planning horizon given the existing capacity of the state.

economic cost estimates can be updated and the planning horizon extended, as necessary, to a period that again does not affect the immediate decision. Note that in the example problem shown in Figures 4.14 and 4.15, the use of 4 periods instead of 5 would have resulted in the same first-period decision. There is no need to extend the analysis to 6 or more periods.

To summarize: What is important to decision-makers now is what additional capacity to add now. While the current period's capacity addition should be based on the best estimates of future costs, interest rates and demands, once a solution is obtained for the capacity expansion required for this and all future periods up to some distant time horizon, one can then ignore all but that first decision, $x_1$: that is, what to add now. Then just before the beginning of the second period, the forecasting and analysis can be redone with updated data to obtain an updated solution for what capacity to add in Period 2, and so on into the future. Thus, these sequential decision-making dynamic programming models can be designed to be used in a sequential decision-making process.

### 4.7.2. Reservoir Operation

Reservoir operators need to know how much water to release and when. Reservoirs designed to meet demands for water supplies, recreation, hydropower, the environment and/or flood control need to be operated in ways that meet those demands in the most reliable and effective manner. Since future inflows or storage volumes are uncertain, the challenge, of course, is to determine the best reservoir release or discharge for a variety of possible inflows and storage conditions.

Reservoir release policies are often defined in the form of what are called 'rule curves'. Figure 4.16 illustrates a rule curve for a single reservoir on the Columbia River in the northwestern United States. It combines components of two basic types of release rules. In both of these, the year is divided into various within-year time periods. There is a specified release for each value of storage in each within-year time period. Usually higher storage zones are associated with higher reservoir releases. If the actual storage is relatively low, then less water is

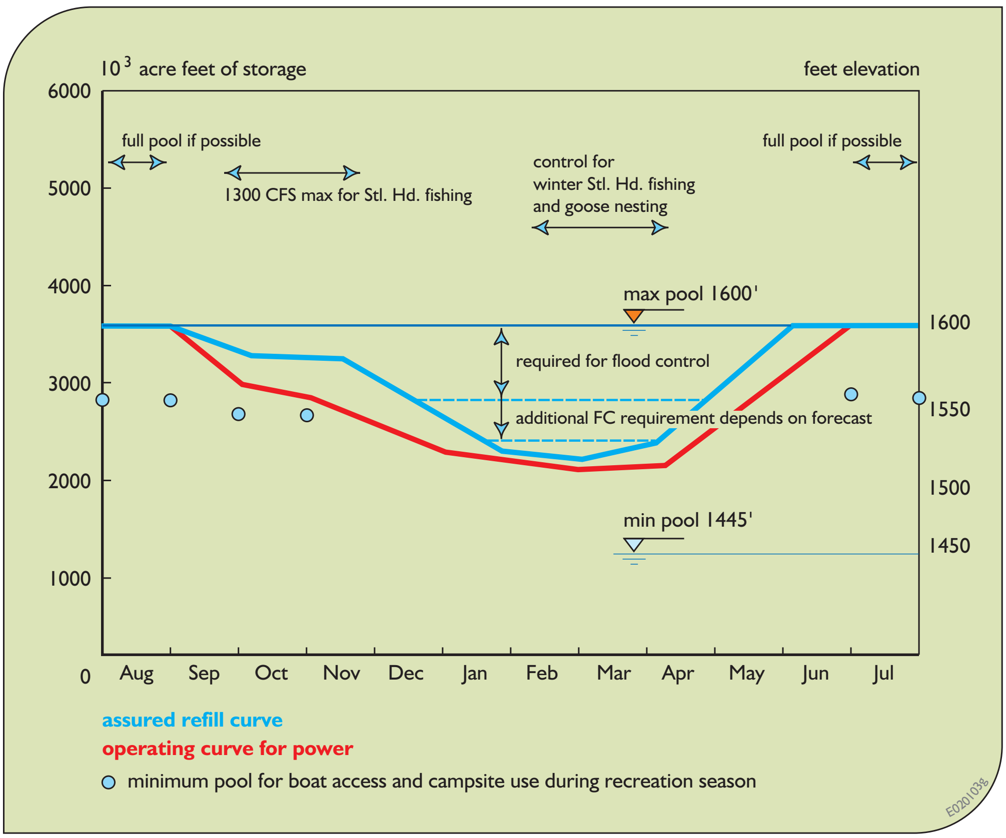



**Figure 4.16.** An example reservoir rule curve specifying the storage targets and some of the release constraints, given the particular current storage volume and time of year. The release constraints also include the minimum and maximum release rates and the maximum downstream channel rate of flow and depth changes that can occur in each month.

usually released so as to hedge against a continuing water shortage or drought.

Release rules may also specify the desired storage level for the time of year. The operator is to release water as necessary to achieve these target storage levels. Maximum and minimum release constraints might also be specified that may affect how quickly the target storage levels can be met. Some rule curves define multiple target storage levels depending on hydrological (e.g., snow pack) conditions in the upstream watershed, or on the forecast climate conditions as affected by ENSO cycles, solar geomagnetic activity, ocean currents and the like. (There is further discussion of this topic in Appendix C).

Reservoir release rule curves for a year, such as that shown in Figure 4.16, define a policy that does not vary from one year to the next. The actual releases will vary, however, depending on the inflows and storage volumes that actually occur. The releases are often specified independently of future inflow forecasts. They are typically based only on existing storage volumes and within-year periods.

Release rules are typically derived from trial and error simulations. To begin these simulations it is useful to have at least an approximate idea of the expected impact of different alternative policies on various system performance measures or objectives. Policy objectives could be the maximization of expected annual net benefits from downstream releases, reservoir storage volumes, hydroelectric energy and flood control, or the minimization of deviations from particular release, storage volume, hydroelectric energy or flood flow targets or target ranges. Discrete dynamic programming can be used to obtain initial estimates of reservoir operating policies that meet these and other objectives. The results of discrete dynamic programming can be expressed in the form shown in Figure 4.16.

### A numerical example

Consider, as a simple example, a reservoir having an active storage capacity of 20 million cubic metres, or for that matter any specified volume units. The active storage volume in the reservoir can vary between 0 and 20. To use discrete dynamic programming, this range of possible storage volumes must be divided into a set of discrete values. These will be the discrete state variable values. In this example let the range of storage volumes be divided into intervals of 5 storage volume units. Hence, the initial storage volume, $S_t$, can assume values of 0, 5, 10, 15 and 20 for all periods $t$.

For each period $t$, let $Q_t$ be the mean inflow, $L_t(S_t, S_{t+1})$ the evaporation and seepage losses that depend on the storage volume in the reservoir, and $R_t$ the release or discharge from the reservoir. Each variable is expressed as volume units for the period $t$.

Storage volume continuity requires that in each period $t$ the initial active storage volume, $S_t$, plus the inflow, $Q_t$, less the losses, $L_t(S_t, S_{t+1})$, and release, $R_t$, equals the final storage, or equivalently the initial storage, $S_{t+1}$, in the following period $t + 1$. Hence:

$$S_t + Q_t - R_t - L_t(S_t, S_{t+1}) = S_{t+1} \quad \text{for each period } t. \tag{4.112}$$

To satisfy the requirement (imposed for convenience in this example) that each storage volume variable be a discrete value ranging from 0 to 20 in units of 5, the releases, $R_t$, must be such that when $Q_t - R_t - L_t(S_t, S_{t+1})$ is added to $S_t$ the resulting value of $S_{t+1}$ is one of the 5 discrete numbers between 0 and 20.

Assume four within-year periods $t$ in each year (kept small for this illustrative example). In these four seasons assume the mean inflows, $Q_t$, are 24, 12, 6 and 18 respectively. Table 4.6 defines the evaporation and seepage losses based on different discrete combinations of initial and final storage volumes for each within-year period $t$.

Rounding these losses to the nearest integer value, Table 4.7 shows the net releases associated with initial and final storage volumes. They are computed using Equation 4.112.

The information in Table 4.7 allows us to draw a network representing each of the discrete storage volume states (the nodes), and each of the feasible releases (the links). This network for the four seasons $t$ in the year is illustrated in Figure 4.17.

This reservoir-operating problem is a multistage decision-making problem. As Figure 4.17 illustrates, at the beginning of any season $t$, the storage volume can be in any of the five discrete states. Given that state, a release decision is to be made. This release will depend on the state: the initial storage volume and the mean inflow, as well as the losses that may be estimated based on the

| period | initial storage | final storage 0 | 5 | 10 | 15 | 20 |
|---|---|---|---|---|---|---|
| period $t = 1$ | **0** | 0 | 0.1 | 0.3 | 0.4 | 0.5 |
| | **5** | 0.1 | 0.2 | 0.4 | 0.6 | 0.8 |
| | **10** | 0.3 | 0.4 | 0.6 | 0.8 | 1.0 |
| | **15** | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 |
| | **20** | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 |
| period $t = 2$ | **0** | 0 | 0.5 | 0.7 | 0.8 | 1.0 |
| | **5** | 0.5 | 0.7 | 0.8 | 1.0 | 1.2 |
| | **10** | 0.7 | 0.8 | 1.0 | 1.2 | 1.4 |
| | **15** | 0.8 | 1.0 | 1.2 | 1.4 | 1.6 |
| | **20** | 1.0 | 1.2 | 1.4 | 1.6 | 1.8 |
| period $t = 3$ | **0** | 0 | 0.7 | 0.9 | 1.0 | 1.2 |
| | **5** | 0.7 | 0.9 | 1.0 | 1.2 | 1.4 |
| | **10** | 0.9 | 1.0 | 1.2 | 1.4 | 1.6 |
| | **15** | 1.0 | 1.2 | 1.4 | 1.6 | 1.8 |
| | **20** | 1.2 | 1.4 | 1.6 | 1.8 | 2.0 |
| period $t = 4$ | **0** | 0 | 0.1 | 0.2 | 0.3 | 0.4 |
| | **5** | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| | **10** | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 |
| | **15** | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| | **20** | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |

losses

E020820n

**Table 4.6.** Evaporation and seepage losses based on initial and final storage volumes for example reservoir operating problem.

| inflow / period | initial storage | final storage 0 | 5 | 10 | 15 | 20 |
|---|---|---|---|---|---|---|
| inflow $Q_1 = 24$, period $t = 1$ | **0** | 24 | 19 | 14 | 8 | 3 |
| | **5** | 29 | 24 | 18 | 13 | 8 |
| | **10** | 34 | 29 | 23 | 18 | 13 |
| | **15** | 39 | 33 | 28 | 23 | 18 |
| | **20** | 43 | 38 | 33 | 28 | 23 |
| inflow $Q_2 = 12$, period $t = 2$ | **0** | 12 | 7 | 1 | – | – |
| | **5** | 17 | 11 | 6 | 1 | – |
| | **10** | 21 | 16 | 11 | 6 | 1 |
| | **15** | 26 | 21 | 16 | 11 | 5 |
| | **20** | 31 | 26 | 21 | 15 | 10 |
| inflow $Q_3 = 6$, period $t = 3$ | **0** | 6 | 0 | – | – | – |
| | **5** | 10 | 5 | 0 | – | – |
| | **10** | 15 | 10 | 5 | 0 | – |
| | **15** | 20 | 15 | 10 | 4 | – |
| | **20** | 25 | 20 | 14 | 9 | 4 |
| inflow $Q_4 = 18$, period $t = 4$ | **0** | 18 | 13 | 8 | 3 | – |
| | **5** | 23 | 18 | 13 | 8 | 3 |
| | **10** | 28 | 23 | 18 | 13 | 7 |
| | **15** | 33 | 28 | 23 | 17 | 12 |
| | **20** | 38 | 33 | 27 | 22 | 17 |

releases

E020820o

**Table 4.7.** Discrete releases associated with initial and final storage volumes.

initial and final storage volumes, as in Table 4.6. The release will also depend on what is to be accomplished – that is, the objectives to be satisfied.

For this example, assume there are various targets that water users would like to achieve. Downstream water users want reservoir operators to meet their flow targets. Individuals who use the lake for recreation want the reservoir operators to meet storage volume or storage level targets. Finally, individuals living on the downstream floodplain want the reservoir operators to provide storage capacity for flood protection. Table 4.8 identifies these different targets that are to be met, if possible, for the duration of each season $t$.

Clearly, it will not be possible to meet all these storage volume and release targets in all four seasons, given inflows of 24, 12, 6 and 18, respectively. Hence, the objective in this example will be to do the best one can: to minimize a weighted sum of squared deviations from each of these targets. The weights reflect the relative importance of meeting each target in each season $t$. Target deviations are squared to reflect the fact that the marginal

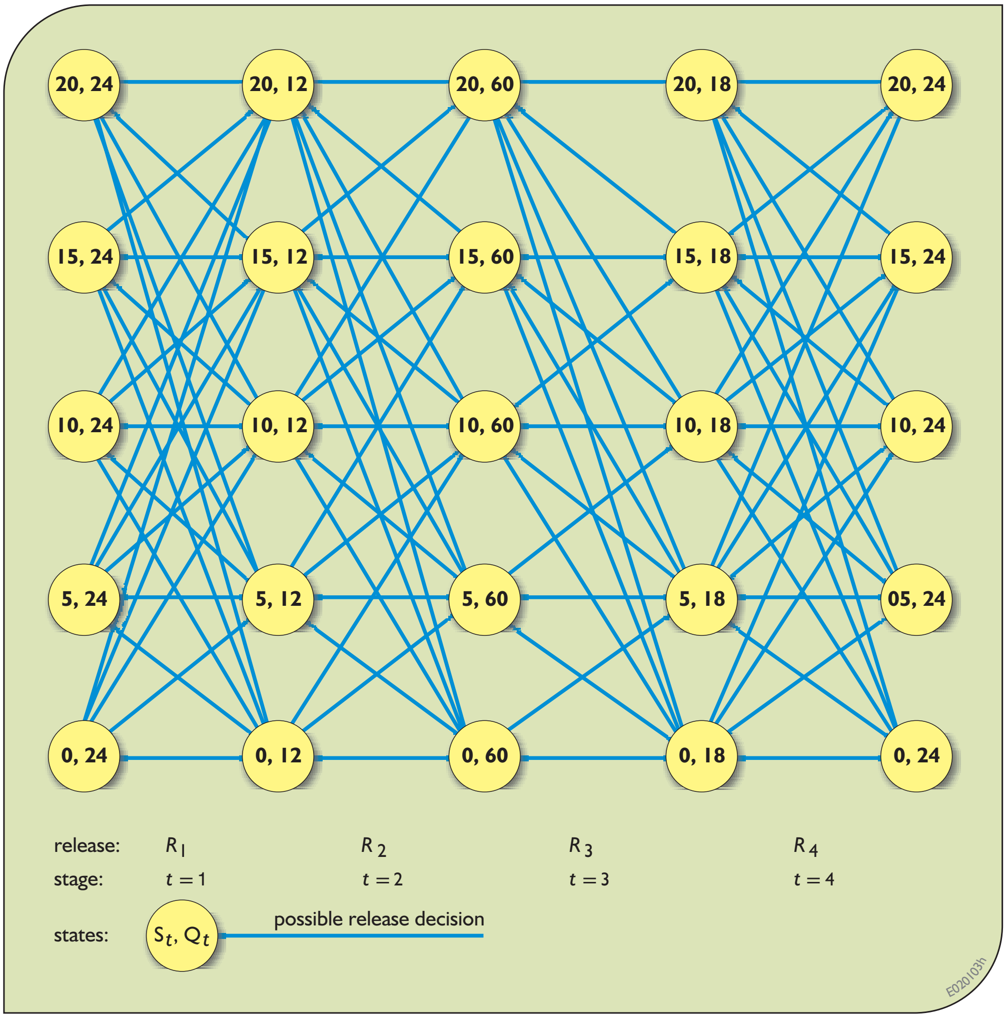



**Figure 4.17.** Network representation of the four-season reservoir release problem. Given any initial storage volume $S_t$ at the beginning of a season $t$, and an expected inflow of $Q_t$ during season $t$, the links indicate the possible release decisions corresponding to those in Table 4.7.

**Table 4.8.** Storage volume and release targets for the example reservoir operation problem.

| period or season t | storage targets $TS^R_t$, $TS^F_t$ | | release target $TR_t$ |
|---|---|---|---|
| 1 | 15 | flood control | ≥ 10 |
| 2 | 20 | recreation | ≥ 15 |
| 3 | 20 | recreation | ≥ 20 |
| 4 | | | ≥ 15 |

E020820p

'losses' associated with deviations increase with increasing deviations. Small deviations are not as serious as larger deviations, and it is better to have numerous small deviations rather than a single larger one.

During the recreation season (Periods 2 and 3), deviations below or above the recreation storage lake volume targets are damaging. During the flood season (Period 1), any storage volume in excess of the flood control storage targets of 15 reduces the flood storage capacity. Deviations below that flood control target are not penalized. Flood control and recreation storage targets during each season $t$ apply throughout the season, thus they apply to the initial storage $S_t$ as well as to the final storage $S_{t+1}$.

The objective is to minimize the sum of total weighted squared deviations, $\mathrm{TSD}_t$, over all seasons $t$ from now on into the future:

$$\text{Minimize } \sum_t \mathrm{TSD}_t \tag{4.113}$$

where

$$\mathrm{TSD}_t = ws_t[(TS_t^R - S_t)^2 + (TS_t^R - S_{t+1})^2] + wfs_t[(ES_t)^2 + (ES_{t+1})^2] + wr_t[DR_t^2] \tag{4.114}$$

In the above equation, when $t = 4$, the last period of the year, $t + 1 = 1$, is the first period in the following year. Each $ES_t$ is the storage volume in excess of the flood storage target volume, $TS_t^F$. Each $DR_t$ is the difference between the actual release, $R_t$, and the target release $TR_t$, when the release is less than the target. The excess storage, $ES_t$, at the beginning of each season $t$ can be defined by the constraint:

$$S_t \leq TS_t^F + ES_t \tag{4.115}$$

for periods $t = 1$ and 2, and the deficit release, $DR_t$, during period $t$ can be defined by the constraint:

$$R_t \geq TR_t - DR_t \tag{4.116}$$

This constraint applies for all periods $t$.

The first component of the right side of Equation 4.114 defines the weighted squared deviations from a recreation storage target, $TS_t^R$, at the beginning and end of season $t$. In this example the recreation season is during Periods 2 and 3. The weights $ws_t$ associated with the recreation component of the objective are 1 in Periods 2 and 3. In Periods 1 and 4 the weights $ws_t$ are 0.

The second component of Equation 4.114 is for flood control. It defines the weighted squared deviations associated with storage volumes in excess of the flood control target volume, $TS_t^F$, at the beginning and end of the flood season, period $t = 1$. In this example, the weights $wfs_t$ are 1 for Period 1 and 0 for Periods 2, 3 and 4. Note the conflict between flood control and recreation at the end of Period 1 or equivalently at the beginning of Period 2.

Finally, the last component of Equation 4.114 defines the weighted squared deficit deviations from a release target, $TR_t$, In this example all release weights $wr_t$ equal 1.

Associated with each link in Figure 4.17 is the release, $R_t$, as defined in Table 4.7. Also associated with each link is the sum of weighted squared deviations, $\mathrm{TSD}_t$, that result from the particular initial and final storage volumes and the storage volume and release targets identified in Table 4.8. They are computed using Equation 4.114, with the releases defined in Table 4.7 and targets defined in Table 4.8, for each feasible combination of initial and final storage volumes, $S_t$ and $S_{t+1}$, for each of the four seasons in a year. These computed weighted squared deviations for each link are shown in Table 4.9.

The goal in this example problem is to find the path through a multi-year network – each year of which is as shown in Figure 4.17 – that minimizes the sum of the squared deviations associated with each of the path's links. Again, each link's weighted squared deviations are given in Table 4.9. Of interest is the best path into the future from any of the nodes or states (discrete storage volumes) that the system could be in at the beginning of any season $t$.

These paths can be found using the backward-moving solution procedure of discrete dynamic programming. This procedure begins at any arbitrarily selected time period or season when the reservoir presumably produces no further benefits to anyone (and it doesn't matter when that time is – just pick any time) and proceeds backward, from right to left one stage (i.e., one time period) at a time, towards the present. At each node (representing a discrete storage volume $S_t$ and inflow $Q_t$), we can calculate the release or final storage volume in that period that minimizes the remaining sum of weighted squared deviations for all remaining seasons. Denote this minimum sum of weighted squared deviations for all $n$ remaining seasons $t$ as $F_t^n(S_t, Q_t)$. This value is dependent on the state $(S_t, Q_t)$, and stage, $t$, and the number $n$ of remaining seasons. It is not a function of the decision $R_t$ or $S_{t+1}$.

This minimum sum of weighted squared deviations for all $n$ remaining seasons $t$ is equal to:

$$F_t^n(S_t, Q_t) = \min \sum_t^n \mathrm{TSD}_t(S_t, R_t, S_{t+1}) \text{ over all feasible values of } R_t \tag{4.117}$$

$TR_1 \geq 10$, $TS_t^F \leq 15$ — period $t = 1$

| initial storage | final storage 0 | 5 | 10 | 15 | 20 |
|---|---|---|---|---|---|
| **0** | 0 | 0 | 0 | 4 | 74 |
| **5** | 0 | 0 | 0 | 0 | 29 |
| **10** | 0 | 0 | 0 | 0 | 25 |
| **15** | 0 | 0 | 0 | 0 | 25 |
| **20** | 25 | 25 | 25 | 25 | 50 |

TSD

$TS_2 = 20$, $TR_2 \geq 15$ — period $t = 2$

| initial storage | final storage 0 | 5 | 10 | 15 | 20 |
|---|---|---|---|---|---|
| **0** | 809 | 689 | 696 | --- | --- |
| **5** | 625 | 466 | 406 | 446 | --- |
| **10** | 500 | 325 | 216 | 206 | 296 |
| **15** | 425 | 250 | 125 | 66 | 125 |
| **20** | 400 | 225 | 100 | 25 | 25 |

TSD

$TS_3 = 20$, $TR_3 \geq 20$ — period $t = 3$

| initial storage | final storage 0 | 5 | 10 | 15 | 20 |
|---|---|---|---|---|---|
| **0** | 996 | 1025 | --- | --- | --- |
| **5** | 725 | 675 | 725 | --- | --- |
| **10** | 525 | 425 | 425 | 525 | --- |
| **15** | 425 | 275 | 225 | 306 | --- |
| **20** | 400 | 225 | 136 | 146 | 256 |

TSD

$TR_4 \geq 15$ — period $t = 4$

| initial storage | final storage 0 | 5 | 10 | 15 | 20 |
|---|---|---|---|---|---|
| **0** | 0 | 4 | 49 | 144 | --- |
| **5** | 0 | 0 | 4 | 49 | 144 |
| **10** | 0 | 0 | 0 | 4 | 64 |
| **15** | 0 | 0 | 0 | 0 | 9 |
| **20** | 0 | 0 | 0 | 0 | 0 |

TSD

E020820q

**Table 4.9. Total sum of squared deviations, $TSD_t$, associated with initial and final storage volumes. These are calculated using Equations 4.114 through 4.116.**

where

$$S_{t+1} = S_t + Q_t - R_t - L_t(S_t, S_{t+1}) \tag{4.118}$$

and

$$S_t \leq K, \text{ the capacity of the reservoir} \tag{4.119}$$

The policy we want to derive is called a *steady-state policy*. Such a policy assumes the reservoir will be operating for a relatively long time with the same objectives. We can find this steady-state policy by first assuming that at some time all future benefits, losses or penalties, $F_t^0(S_t, Q_t)$, will be 0. We can begin in that season $t$ and work backwards towards the present, moving left through the network one season $t$ at a time. We can continue for multiple years until the annual policy begins repeating itself each year. In other words, when the optimal $R_t$ associated with a particular state $(S_t, Q_t)$ is the same in two or more successive years, and this applies for all states $(S_t, Q_t)$ in each season $t$, a steady-state policy has probably been obtained (a more definitive test of whether or not a steady-state policy has been reached will be discussed later). A steady-state policy will occur if the inflows, $Q_t$, and objectives, $TSD_t(S_t, R_t, S_{t+1})$, remain the same from year to year. This steady-state policy is independent of the assumption that the operation will end at some point.

To find the steady-state operating policy for this example problem, assume the operation ends in some distant year at the end of Season 4 (the right-hand side nodes in Figure 4.17). At the end of this season the number of remaining seasons, $n$, equals 0. The values of the remaining minimum sums of weighted squared deviations, $F_1^0(S_1, Q_1)$ associated with each state $(S_1, Q_1)$, i.e., each node, equal 0. Now we can begin the process of finding the best releases $R_t$ in each successive season, moving backward to the beginning of stage $t = 4$, then stage $t = 3$, then to $t = 2$, and then to $t = 1$, and then to $t = 4$ of the preceding year, and so on, each move to the left increasing the number $n$ remaining seasons by one.

At each stage, or season, we can compute the release $R_t$ or equivalently the final storage volume $S_{t+1}$, that minimizes

$$F_t^n(S_t, Q_t) = \text{Minimum}\{TSD_t(S_t, R_t, S_{t+1}) + F_{t+1}^{n-1}(S_{t+1}, Q_{t+1})\} \quad \text{for all } 0 \leq S_t \leq 20 \tag{4.120}$$

The decision-variable can be either the release, $R_t$, or the final storage volume, $S_{t+1}$. If the decision-variable is the release, then the constraints on that release $R_t$ are:

$$R_t \leq S_t + Q_t - L_t(S_t, S_{t+1}) \tag{4.121}$$

$$R_t \geq S_t + Q_t - L_t(S_t, S_{t+1}) - 20 \text{ (the capacity)} \tag{4.122}$$

and

$$S_{t+1} = S_t + Q_t - R_t - L_t(S_t, S_{t+1}) \tag{4.123}$$

If the decision-variable is the final storage volume, the constraints on that final storage volume $S_{t+1}$ are:

$$0 \leq S_{t+1} \leq 20 \tag{4.124}$$

$$S_{t+1} \leq S_t + Q_t - L_t(S_t, S_{t+1}) \tag{4.125}$$

and

$$R_t = S_t + Q_t - S_{t+1} - L_t(S_t, S_{t+1}) \tag{4.126}$$

Note that if the decision-variable is $S_{t+1}$ in season $t$, this decision becomes the state variable in season $t+1$. In both cases, the storage volumes in each season are limited to discrete values 0, 5, 10, 15 and 20.

Tables 4.10 through 4.19 show the values obtained from solving the recursive equations for 10 successive seasons or stages (2.5 years). Each table represents a stage or season $t$, beginning with Table 4.10 at $t = 4$ and the number of remaining seasons $n = 1$. The data in each table are obtained from Tables 4.7 and 4.9. The last two columns of each table represent the best release and final storage volume decision(s) associated with the state (initial storage volume and inflow).

Note that the policy defining the release or final storage for each discrete initial storage volume in season $t = 3$ in Table 4.12 is the same as in Table 4.16, and similarly for season $t = 4$ in Tables 4.13 and 4.17, and for season $t = 1$ in Tables 4.14 and 4.18, and finally for season $t = 2$ in Tables 4.15 and 4.19. The policy differs over each state, and over each different season, but not from year to year for any specified state and season. We have reached a steady-state policy. If we kept on computing the release and final storage policies for preceding seasons, we would

**Table 4.10.** Calculation of minimum squared deviations associated with various discrete storage states in season $t = 4$ with only $n = 1$ season remaining for reservoir operation.

period $t = 4$ $n = 1$ $Q_4 = 18$

$F_4^1(S_4, Q_4) = \min TSD_4$

| initial storage $S_4$ | $TSD_4$ final storage $S_1$: 0 | 5 | 10 | 15 | 20 | $F_4^1$ | $R_4$ | $S_1$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 4 | 49 | 144 | - | 0 | 18 | 0 |
| 5 | 0 | 0 | 4 | 49 | 144 | 0 | 18-23 | 0-5 |
| 10 | 0 | 0 | 0 | 4 | 64 | 0 | 18-28 | 0-10 |
| 15 | 0 | 0 | 0 | 0 | 9 | 0 | 17-33 | 0-15 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 17-38 | 0-20 |

E020827a

**Table 4.11.** Calculation of minimum squared deviations associated with various discrete storage states in season $t = 3$ with $n = 2$ seasons remaining for reservoir operation.

period $t = 3$ $n = 2$ $Q_3 = 6$

$F_3^2(S_3, Q_3) = \min \{TSD_3(S_3, R_3, S_4) + F_4^1(S_4, Q_4)\}$

| initial storage $S_3$ | $TSD_3 + F_4^1(S_4, Q_4)$ final storage $S_4$: 0 | 5 | 10 | 15 | 20 | $F_3^2$ | $R_3$ | $S_4$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 996 | 1025 | - | - | - | 996 | 6 | 0 |
| 5 | 725 | 675 | 725 | - | - | 675 | 5 | 5 |
| 10 | 525 | 425 | 425 | 525 | - | 425 | 5-10 | 5-10 |
| 15 | 425 | 275 | 225 | 306 | - | 225 | 10 | 10 |
| 20 | 400 | 225 | 136 | 146 | 256 | 136 | 14 | 10 |

E020827b

period t = 2 $n$ = 3 $Q_2$ = 12

$F_2^3(S_2, Q_2) = \min\{TSD_2(S_2, R_2, S_3) + F_3^2(S_3, Q_3)\}$

| initial storage $S_2$ | $TSD_2 + F_3^2(S_3, Q_3)$ final storage $S_3$: 0 | 5 | 10 | 15 | 20 | $F_2^3$ | $R_2$ | $S_3$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 809+996 | 689+675 | 696+425 | - | - | 1121 | 1 | 10 |
| 5 | 625+996 | 466+675 | 406+425 | 446+225 | - | 671 | 1 | 15 |
| 10 | 500+996 | 325+675 | 216+425 | 206+225 | 296+136 | 431 | 6 | 15 |
| 15 | 425+996 | 250+675 | 125+425 | 66+225 | 125+136 | 261 | 5 | 20 |
| 20 | 400+996 | 225+675 | 100+425 | 25+225 | 25+136 | 161 | 10 | 20 |

E020827c

**Table 4.12.** Calculation of minimum squared deviations associated with various discrete storage states in season $t = 2$ with $n = 3$ seasons remaining for reservoir operation.

period t = 1 $n$ = 4 $Q_1$ = 24

$F_1^4(S_1, Q_1) = \min\{TSD_1(S_1, R_1, S_2) + F_2^3(S_2, Q_2)\}$

| initial storage $S_1$ | $TSD_1 + F_2^3(S_2, Q_2)$ final storage $S_2$: 0 | 5 | 10 | 15 | 20 | $F_1^4$ | $R_1$ | $S_2$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0+1121 | 0+671 | 0+431 | 4+261 | 74+161 | 235 | 3 | 20 |
| 5 | 0+1121 | 0+671 | 0+431 | 0+261 | 29+161 | 190 | 8 | 20 |
| 10 | 0+1121 | 0+671 | 0+431 | 0+261 | 25+161 | 186 | 13 | 20 |
| 15 | 0+1121 | 0+671 | 0+431 | 0+261 | 25+161 | 186 | 18 | 20 |
| 20 | 25+1121 | 25+671 | 25+431 | 25+261 | 50+161 | 211 | 23 | 20 |

E020827d

**Table 4.13.** Calculation of minimum squared deviations associated with various discrete storage states in season $t = 1$ with $n = 4$ seasons remaining for reservoir operation.

period t = 4 $n$ = 5 $Q_4$ = 18

$F_4^5(S_4, Q_4) = \min\{TSD_4(S_4, R_4, S_1) + F_1^4(S_1, Q_1)\}$

| initial storage $S_4$ | $TSD_4 + F_1^4(S_1, Q_1)$ final storage $S_1$: 0 | 5 | 10 | 15 | 20 | $F_4^5$ | $R_4$ | $S_1$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0+235 | 4+190 | 49+186 | 144+186 | - | 194 | 13 | 5 |
| 5 | 0+235 | 0+190 | 4+186 | 49+186 | 144+211 | 190 | 13-18 | 5-10 |
| 10 | 0+235 | 0+190 | 0+186 | 4+186 | 64+211 | 186 | 18 | 10 |
| 15 | 0+235 | 0+190 | 0+186 | 0+186 | 9+211 | 186 | 17-23 | 10-15 |
| 20 | 0+235 | 0+190 | 0+186 | 0+186 | 0+211 | 186 | 22-27 | 10-15 |

E020827e

**Table 4.14.** Calculation of minimum squared deviations associated with various discrete storage states in season $t = 4$ with $n = 5$ seasons remaining for reservoir operation.

**Table 4.15.** Calculation of minimum squared deviations associated with various discrete storage states in season $t = 3$ with $n = 6$ seasons remaining for reservoir operation.

period t = 3 $n = 6$ $Q_3 = 6$

$F_3^6(S_3, Q_3) = \min\{TSD_3(S_3, R_3, S_4) + F_4^5(S_4, Q_4)\}$

| initial storage $S_3$ | $TSD_3 + F_4^5(S_4, Q_4)$ final storage $S_4$: 0 | 5 | 10 | 15 | 20 | $F_3^6$ | $R_3$ | $S_4$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 996+194 | 1025+190 | - | - | - | 1190 | 6 | 0 |
| 5 | 725+194 | 675+190 | 725+186 | - | - | 865 | 5 | 5 |
| 10 | 525+194 | 425+190 | 425+186 | 525+186 | - | 611 | 5 | 10 |
| 15 | 425+194 | 275+190 | 225+186 | 306+186 | - | 411 | 10 | 10 |
| 20 | 400+194 | 225+190 | 136+186 | 146+186 | 256+186 | 322 | 14 | 10 |

E020827f

**Table 4.16.** Calculation of minimum squared deviations associated with various discrete storage states in season $t = 2$ with $n = 7$ seasons remaining for reservoir operation.

period t = 2 $n = 7$ $Q_2 = 12$

$F_2^7(S_2, Q_2) = \min\{TSD_2(S_2, R_2, S_3) + F_3^6(S_3, Q_3)\}$

| initial storage $S_2$ | $TSD_2 + F_3^6(S_3, Q_3)$ final storage $S_3$: 0 | 5 | 10 | 15 | 20 | $F_2^7$ | $R_2$ | $S_3$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 809+1190 | 689+865 | 696+611 | - | - | 1307 | 1 | 10 |
| 5 | 625+1190 | 466+865 | 406+611 | 446+411 | - | 857 | 1 | 15 |
| 10 | 500+1190 | 325+865 | 216+611 | 206+411 | 296+322 | 617 | 6 | 15 |
| 15 | 425+1190 | 250+865 | 125+611 | 66+411 | 125+322 | 447 | 5 | 20 |
| 20 | 400+1190 | 225+865 | 100+611 | 25+411 | 25+322 | 347 | 10 | 20 |

E020827g

**Table 4.17.** Calculation of minimum squared deviations associated with various discrete storage states in season $t = 1$ with $n = 8$ seasons remaining for reservoir operation.

period t = 1 $n = 8$ $Q_2 = 24$

$F_1^8(S_1, Q_1) = \min\{TSD_1(S_1, R_1, S_2) + F_2^7(S_2, Q_2)\}$

| initial storage $S_1$ | $TSD_1 + F_2^7(S_2, Q_2)$ final storage $S_2$: 0 | 5 | 10 | 15 | 20 | $F_1^8$ | $R_1$ | $S_2$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0+1307 | 0+857 | 0+617 | 4+447 | 74+347 | 421 | 3 | 20 |
| 5 | 0+1307 | 0+857 | 0+617 | 0+447 | 29+347 | 376 | 8 | 20 |
| 10 | 0+1307 | 0+857 | 0+617 | 0+447 | 25+347 | 372 | 13 | 20 |
| 15 | 0+1307 | 0+857 | 0+617 | 0+447 | 25+347 | 372 | 18 | 20 |
| 20 | 25+1307 | 25+857 | 25+617 | 25+447 | 50+347 | 397 | 23 | 20 |

E020827h

period t = 4  $n = 9$  $Q_4 = 18$

$F_4^9(S_4, Q_4) = \min\{TSD_4(S_4, R_4, S_1) + F_1^8(S_1, Q_1)\}$

| initial storage $S_4$ | $TSD_4 + F_1^8(S_1, Q_1)$ final storage $S_1$: 0 | 5 | 10 | 15 | 20 | $F_4^9$ | $R_4$ | $S_1$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0+421 | 4+376 | 49+372 | 144+372 | - | 380 | 13 | 5 |
| 5 | 0+421 | 0+376 | 4+372 | 49+372 | 144+397 | 376 | 13-18 | 5-10 |
| 10 | 0+421 | 0+376 | 0+372 | 4+372 | 64+397 | 372 | 18 | 10 |
| 15 | 0+421 | 0+376 | 0+372 | 0+372 | 9+397 | 372 | 17-23 | 10-15 |
| 20 | 0+421 | 0+376 | 0+372 | 0+372 | 0+397 | 372 | 22-27 | 10-15 |

E020827k

**Table 4.18.** Calculation of minimum squared deviations associated with various discrete storage states in season $t = 4$ with $n = 9$ seasons remaining for reservoir operation.

period t = 3  $n = 10$  $Q_3 = 6$

$F_3^{10}(S_3, Q_3) = \min\{TSD_3(S_3, R_3, S_4) + F_4^9(S_4, Q_4)\}$

| initial storage $S_3$ | $TSD_3 + F_4^9(S_4, Q_4)$ final storage $S_4$: 0 | 5 | 10 | 15 | 20 | $F_3^{10}$ | $R_3$ | $S_4$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 996+380 | 1025+376 | - | - | - | 1376 | 6 | 0 |
| 5 | 725+380 | 675+376 | 725+372 | - | - | 1051 | 5 | 5 |
| 10 | 525+380 | 425+376 | 425+372 | 525+372 | - | 797 | 5 | 10 |
| 15 | 425+380 | 275+376 | 225+372 | 306+372 | - | 597 | 10 | 10 |
| 20 | 400+380 | 225+376 | 136+372 | 146+372 | 256+372 | 508 | 14 | 10 |

E020827m

**Table 4.19.** Calculation of minimum squared deviations associated with various discrete storage states in season $t = 3$ with $n = 10$ seasons remaining for reservoir operation.

get the same policy as that found for the same season in the following year. The policy is dependent on the state – the initial storage volume in this case – and on the season $t$, but not on the year. This policy as defined in Tables 4.16–4.19 is summarized in Table 4.20.

This policy can be defined as a rule curve, as shown in Figure 4.18. It provides a first approximation of a reservoir release rule curve that one can improve upon using simulation.

Table 4.20 and Figure 4.18 define a policy that can be implemented for any initial storage volume condition at the beginning of any season $t$. This can be simulated under different flow patterns to determine just how well it satisfies the overall objective of minimizing the weighted sum of squared deviations from desired, but conflicting, storage and release targets. There are other performance criteria that may also be evaluated using simulation, such as measures of reliability, resilience, and vulnerability (Chapter 10).

Assuming the inflows that were used to derive this policy actually occurred each year, we can simulate the derived sequential steady-state policy to find the storage volumes and releases that would occur in each period, year after year, once a repetitive steady-state condition were reached. This is done in Table 4.21 for an arbitrary initial storage volume of 20 in season $t = 1$. You can try other initial conditions to verify that it requires only two years at most to reach a repetitive steady-state policy.

As shown in Table 4.21, if the inflows were repetitive and the optimal policy was followed, the initial storage

**Table 4.20.** The discrete steady-state reservoir operating policy as computed for this example problem in Tables 4.16 to 4.19.

| initial storage S | release season t: 1 | 2 | 3 | 4 | final storage volume season t: 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|
| **0** | 3 | 1 | 6 | 13 | 20 | 10 | 0 | 5 |
| **5** | 8 | 1 | 5 | 13-18 | 20 | 15 | 5 | 5-10 |
| **10** | 13 | 6 | 5 | 18 | 15 | 10 | 10 | 10 |
| **15** | 18 | 5 | 10 | 17-23 | 20 | 20 | 10 | 10-15 |
| **20** | 23 | 10 | 14 | 22-27 | 20 | 20 | 10 | 10-15 |

E020827n

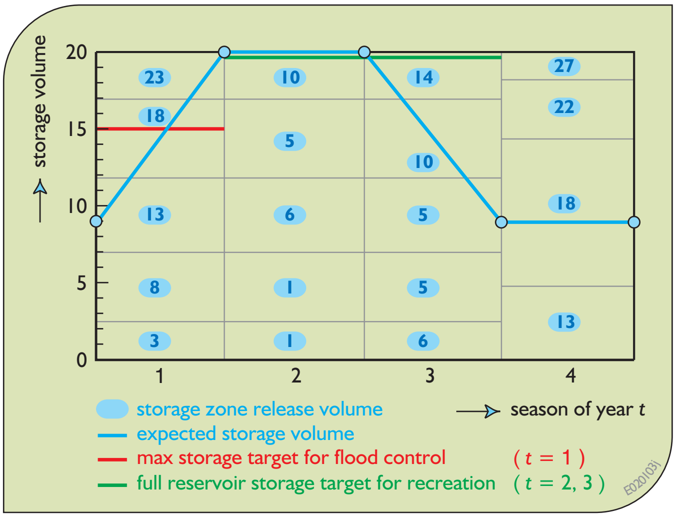



**Figure 4.18.** Reservoir rule curve based on policy defined in Table 4.20. Each season is divided into storage volume zones. The releases associated with each storage volume zone are specified. Also shown are the storage volumes that would result if in each year the actual inflows equalled the inflows used to derive this rule curve.

volumes and releases would begin to repeat themselves once a steady-state condition has been reached. Once reached, the storage volumes and releases will be the same each year (since the inflows are the same). These storage volumes are denoted as a blue line on the rule curve shown in Figure 4.18. The annual total squared deviations will also be the same each year. As seen in Table 4.21, this annual minimum weighted sum of squared deviations for this example equals 186. This is what would be observed if the inflows assumed for this analysis repeated themselves.

Note from Tables 4.12 – 4.15 and 4.16 – 4.19 that once the steady-state sequential policy has been reached for any specified storage volume, $S_t$, and season $t$, the annual difference of the accumulated minimum sum of squared deviations, $F_t^n(S_t, Q_t)$, equals a constant, namely the annual value of the objective function. In this case that constant is 186.

$$F_t^{n+4}(S_t, Q_t) - F_t^n(S_t, Q_t) = 186 \quad \text{for all } S_t, Q_t \text{ and } t. \tag{4.127}$$

This condition indicates a steady-state policy has been achieved.

This policy in Table 4.21 applies only for the assumed inflows in each season. It does not define what to do if the initial storage volumes or inflows differ from those for which the policy is defined. Initial storage volumes and inflows can and will vary from those specified in the solution of any deterministic model. One fact is certain: no matter what inflows are assumed in any model, the actual inflows will always differ. Hence, a policy as defined in Table 4.20 and Figure 4.18 is much more useful than that in Table 4.21. But as for any output from a relatively simple optimization model, this policy should be simulated, evaluated and further refined in an effort to identify the policy that best meets the operating policy objectives.

## 4.8. General Comments on Dynamic Programming

Before ending this discussion of dynamic programming methods and its applicability for analysing water resources planning problems, we should examine a major

| year | season $t$ | $S_t$ + | $Q_t$ − | $R_t$ − | $L_t$ = | $S_{t+1}$ | $TSD_t$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 20 | 24 | 23 | 1 | 20 | |
| 1 | 2 | 20 | 12 | 10 | 2 | 20 | |
| 1 | 3 | 20 | 6 | 14 | 2 | 10 | |
| 1 | 4 | 10 | 18 | 18 | 0 | 10 | |

| year | season $t$ | $S_t$ + | $Q_t$ − | $R_t$ − | $L_t$ = | $S_{t+1}$ | $TSD_t$ |
|---|---|---|---|---|---|---|---|
| 2 | 1 | 10 | 24 | 14 | 0 | 20 | 25 |
| 2 | 2 | 20 | 12 | 10 | 2 | 20 | 25 |
| 2 | 3 | 20 | 6 | 14 | 2 | 10 | 136 |
| 2 | 4 | 10 | 18 | 18 | 0 | 10 | 0 |

| year | season $t$ | $S_t$ + | $Q_t$ − | $R_t$ − | $L_t$ = | $S_{t+1}$ | $TSD_t$ |
|---|---|---|---|---|---|---|---|
| 3 | 1 | 10 | etc ... | repeating ... | | | |
| | | total sum of squared deviations, **TSD**, = | | | | | 186 |

E020827p

**Table 4.21.** A simulation of the derived operating policy in Table 4.20. The storage volumes and releases in each period $t$ will repeat themselves each year, after the first year. The annual total squared deviations, $TSD_t$ for the specific initial and final storage volumes and release conditions are obtained from Table 4.9.

assumption that has been made in each of the applications presented. The first is that the net benefits or costs or other objective values resulting from each decision at each stage of the problem are dependent only on the state variable values in each stage. They are independent of decisions made at other stages. If the returns at any stage are dependent on the decisions made at other stages in a way not captured by the state variables. For example, if the release and storage targets or the reservoir capacity were unknown in the previous reservoir operating policy problem, then dynamic programming, with some exceptions, becomes much more difficult to apply. Dynamic programming models can be applied to design problems, such as the capacity-expansion problem or a reservoir storage capacity–yield relationship as will be discussed later, or to operating problems, such as the water allocation and reservoir operation problems, but rarely to problems having both unknown design and operating policy decision-variables. While there are some tricks that may allow dynamic programming to be used to find the best solutions to both design and operating problems encountered in water resources planning and management studies, other optimization methods, perhaps combined with dynamic programming where appropriate, are often more useful.

## 5. Linear Programming

If the objective function and constraints of an optimization model are all linear, there are many readily available computer programs one can use to find their solutions. These programs are very powerful, and unlike many other optimization methods, they can be applied successfully to very large optimization problems. Many water resources problems contain many variables and constraints, too many to be easily solved using non-linear or dynamic programming methods. Linear programming procedures or algorithms for solving linear optimization models are often the most efficient ways to find solutions to such problems.

Because of the availability of computer programs that can solve linear programming problems, linear programming is arguably the most popular and commonly applied optimization algorithm. It is used to identify and evaluate alternative plans, designs and management policies in agriculture, business, commerce, education, engineering, finance, the civil and military branches of government, and many other fields.

Many models of complex water resources systems are, or can be made, linear. Many are also very large. The number of variables and constraints simply defining mass balances and capacity limitations alone at many river basin sites and for numerous time periods can become so big as to preclude the practical use of most other optimization methods. Because of the power and availability of computer programs that can solve large linear programming problems, a variety of methods have been developed to approximate non-linear (especially separable) functions with linear ones just so linear programming can be used to solve various otherwise non-linear problems. Some of these methods will be described shortly.

In spite of its power and popularity, for most real-world water resources planning and management problems, linear programming, like the other optimization methods already discussed in this chapter, is best viewed as a preliminary screening tool. Its value is more for reducing the number of alternatives for further more detailed simulations than for finding the best decision. This is not just because approximation methods may have been used to convert non-linear functions to linear ones, but more likely because it is difficult to incorporate all the complexity of the system and all the objectives considered important to all stakeholders into the linear model. Nevertheless, linear programming, like other optimization methods, can provide initial designs and operating policy information that simulation models require before they can simulate those designs and operating policies.

Equations 4.45 and 4.46 define the general structure of any constrained optimization problem. If the objective function $F(\boldsymbol{X})$ of the vector $\boldsymbol{X}$ of decision-variables $x_j$ is linear and if all the constraints $g_i(\boldsymbol{X})$ in Equation 4.46 are linear, then the model becomes a linear programming model. The general structure of a linear programming model is:

$$\text{Maximize or minimize } \sum_{j}^{n} P_j x_j \tag{4.128}$$

$$\text{Subject to: } \sum_{j}^{n} a_{ij} x_j \le b_i \quad \text{for } i = 1, 2, 3, \ldots, m \tag{4.129}$$

$$x_j \ge 0 \quad \text{for all } j = 1, 2, 3, \ldots, n. \tag{4.130}$$

If any model fits this general form, where the constraints can be any combination of equalities (=) and inequalities (≥ or ≤), then a large variety of linear programming computer programs can be used to find the 'optimal' values of all the unknown decision-variables $x_j$. With some exceptions, variable non-negativity is enforced within the solution algorithms of most commercial linear programming programs, eliminating the need to have to specify these conditions in any particular application.

Users of linear programming need to know how to construct linear models and how to use the computer programs that are available for solving them. They do not have to understand all the mathematical details of the solution procedure incorporated in the linear programming codes. But users of linear programming computer programs should understand what the solution procedure does and what the computer program output means. To begin this discussion of these topics, consider some simple examples of linear programming models.

## 5.1. Reservoir Storage Capacity–Yield Models

Linear programming can be used to define storage capacity–yield functions for a single or multiple reservoirs. A storage capacity–yield function defines the maximum constant 'dependable' reservoir release or yield that will be available, at a given level of reliability, during each period of operation, as a function of the active storage volume capacity. The yield from any reservoir or group of reservoirs will depend on the active storage capacity of each reservoir and the water that flows into each reservoir, its inflow. Figure 4.19 illustrates two typical storage–yield functions for a single reservoir.

To describe what a yield is and how it can be increased, consider a sequence of 5 annual flows, say 2, 4, 1, 5 and 3, at a site in an unregulated stream. Based on this admittedly very limited record of flows, the minimum (historically) 'dependable' annual flow yield of the stream at that site is 1, the minimum observed flow. Assuming the flow record is representative of what future flows might be, a discharge of 1 can be 'guaranteed' in each of non-zero the five time-periods of record. (In reality, that or any non-zero yield will have a reliability less than 1, as will be considered in Chapter 11).

If a reservoir having an active storage capacity of 1 is built, it could store 1 volume unit of flow when the flow is equal to or greater than 2, and then release it along with the natural flow when the natural flow is 1, increasing the

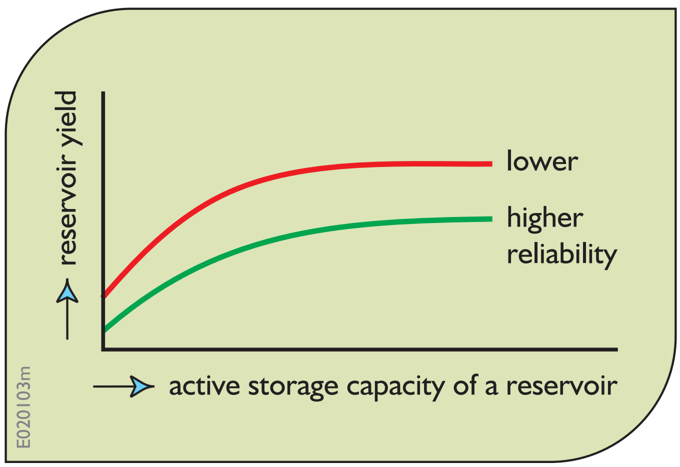



**Figure 4.19.** Two storage–yield functions for a single reservoir defining the maximum minimum dependable release. These functions can be defined for varying levels of yield reliability.

minimum dependable flow to 2 units in each year. Storing 2 units when the flow is 5, releasing 1 and the natural flow when that natural flow is 2, and storing 1 when the flow is 4, and then releasing the stored 2 units along with the natural flow when the natural flow is 1, will permit a yield of 3 in each time period with 2 units of active capacity. This is the maximum annual yield that is possible at this site, again based on these five annual inflows and their sequence. The maximum annual yield cannot exceed the mean annual flow, which in this example is 3. Hence, the storage capacity–yield function equals 1 when the active capacity is 0, 2 when the active capacity is 1, and 3 when the active capacity is 2. The annual yield remains at 3 for any active storage capacity in excess of 2.

This storage–yield function is dependent not only on the natural unregulated annual flows but also on their sequence. For example if the sequence of the same 5 annual flows were 5, 2, 1, 3, 4, the needed active storage capacity is 3 instead of 2 volume units as before to obtain a dependable flow or yield of 3 volume units. In spite of these limitations of storage capacity–yield functions, historical records are still typically used to derive them. (Ways of augmenting the historical flow record are discussed in Chapter 7.)

There are many methods available for deriving storage–yield functions. One very versatile method, especially for multiple reservoir systems, uses linear programming. Others are discussed in Chapter 11.

To illustrate a storage capacity–yield model, consider a single reservoir that must provide a minimum release or yield $Y$ in each period $t$. Assume a record of known (historical or synthetic) streamflows at the reservoir site is available. The problem is to find the maximum uniform yield $Y$ obtainable from a given active storage capacity. The objective is to

$$\text{maximize } Y \tag{4.131}$$

This maximum yield is constrained by the water available in each period, and by the reservoir capacity. Two sets of constraints are needed to define the relationships among the inflows, the reservoir storage volumes, the yields, any excess release, and the reservoir capacity. The first set of continuity equations equate the unknown final reservoir storage volume $S_{t+1}$ in period $t$ to the unknown initial reservoir storage volume $S_t$ plus the known inflow $Q_t$, minus the unknown yield $Y$ and excess release, $R_t$, if any, in period $t$. (Losses are being ignored in this example.)

$$S_t + Q_t - Y - R_t = S_{t+1}$$
$$\text{for each period } t = 1, 2, 3, \dots, T. \quad T+1 = 1 \tag{4.132}$$

If, as indicated in Equation 4.132, one assumes that Period 1 follows the last Period $T$, it is not necessary to specify the value of the initial storage volume $S_1$ and/or final storage volume $S_{T+1}$. They are set equal to each other and that variable value remains unknown. The resulting 'steady-state' solution is based on the inflow sequence that is assumed to repeat itself.

The second set of required constraints ensures that the reservoir storage volumes $S_t$ at the beginning of each period $t$ are no greater than the active reservoir capacity $K$.

$$S_t \leq K \quad t = 1, 2, 3, \dots, T \tag{4.133}$$

To derive a storage–yield function, the model defined by Equations 4.131, 4.132 and 4.133 must be solved for various assumed values of capacity $K$. Only the inflow values $Q_t$ and reservoir active storage capacity $K$ are known. All other storage, release and yield variables are unknown. Linear programming will be able to find their optimal values. Clearly, the upper bound on the yield regardless of reservoir capacity will equal the mean inflow (less any losses if they were included).

Alternatively, one can solve a number of linear programming models that minimize an unknown storage capacity $K$ needed to achieve various specified yields $Y$. The resulting storage–yield function will be same; the minimum capacity needed to achieve a specified yield will be the same as the maximum yield obtainable from the corresponding specified capacity $K$. However, the specified yield $Y$ cannot exceed the mean inflow. If it does, there will be no feasible solution to the linear programming model.

Box 4.1 illustrates an example storage–yield model and its solution to find the storage–yield function. For this problem, and others in this chapter, the program LINGO (www.lindo.com) is used.

Before moving to another application of linear programming, consider how this storage–yield problem, Equations 4.131–4.133, can be formulated as a discrete dynamic programming model. The use of discrete dynamic programming is clearly not the most efficient way to define a storage–yield function but the problem of

### Box 4.1. Example storage capacity–yield model and its solution from LINGO

E020903a

```
! Reservoir Storage-Yield Model:
Define St as the initial active res. storage, period t,
Y as the reliable yield in each period t,
Rt as the excess release from the res., period t,
Qt as the known inflow volume to the res., period t
K as the reservoir active storage volume capacity.
;
Max = Y ; !Applies to Model 1. Must be omitted for Model 2;
Min = K ; !Applies to Model 2. Must be omitted for Model 1;
!
Subject to:
Mass balance constraints for each of 5 periods t.
;
S1 + Q1 - Y - R1 = S2;
S2 + Q2 - Y - R2 = S3;
S3 + Q3 - Y - R3 = S4;
S4 + Q4 - Y - R4 = S5;
S5 + Q5 - Y - R5 = S1; ! assumes a steady -state condition;
!
Capacity constraints on storage volumes.
;
S1 < K; S2 < K; S3 < K; S4 < K; S5 < K;
Data:
Q1 = 10; Q2 = 5; Q3 = 30; Q4 = 20; Q5 = 15;
!Note mean = 16;
K = ? ; ! Use for Model 1 only. Allows user to enter
any value of K during model run.;
Y = ? ; ! Use for Model 2 only. Allows user to enter
any value of Y during model run.
Enddata
```

**model solutions for specified values of *K* in model 1 or values of *Y* in model 2**

| K | Y | S1 | S2 | S3 | S4 | S5 | R1 | R2 | R3 | R4 | R5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 5 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 25 | 15 | 10 |
| 5 | 10 | 5 | 5 | 0 | 5 | 5 | 0 | 0 | 15 | 10 | 5 |
| 10 | 12.5 | 10 | 7.5 | 0 | 2.5 | 10 | 0 | 0 | 15 | 0 | 2.5 |
| 15 | 15 | 10 | 10 | 0 | 15 | 15 | 0 | 0 | 0 | 5 | 0 |
| 18 | 16 | 17 | 11 | 0 | 14 | 18 | 0 | 0 | 0 | 0 | 0 |

finding a storage–yield function provides a good exercise in dynamic programming. The dynamic programming network has the same form as shown in Figure 4.19, where each node is a discrete storage and inflow state, and the links represent releases. Let $F_t^n(S_t)$ be the maximum yield obtained given a storage volume of $S_t$ in period $t$ of a year with $n$ periods remaining of reservoir operation. For initial conditions, assume all values of $F_t^0(S_t)$ for some final period $t$ with no more periods $n$ remaining equal a large number that exceeds the mean annual inflow. Then for the set of feasible discrete total releases $R_t$:

$$F_t^n(S_t) = \max\{\min[R_t, F_{t+1}^{n-1}(S_{t+1})]\} \tag{4.134}$$

This applies for all discrete storage volumes $S_t$ and for all within-year periods $t$ and remaining periods $n$. The constraints on the decision-variables $R_t$ are:

$$R_t \leq S_t + Q_t,$$

$$R_t \geq S_t + Q_t - K, \quad \text{and}$$

$$S_{t+1} = S_t + Q_t - R_t \tag{4.135}$$

These recursive Equations 4.134 together with constraint Equations 4.135 can be solved, beginning with $n = 1$ and then for successive values of seasons $t$ and remaining periods $n$, until a steady-state solution is obtained, that is, until

$$F_t^n(S_t) - F_t^{n-T}(S_t) = 0 \quad \text{for all values of } S_t \text{ and periods } t. \tag{4.136}$$

The steady-state yields $F_t(S_t)$ will depend on the storage volumes $S_t$. High initial storage volumes will result in higher yields than will lower ones. The highest yield will be that associated with the highest storage volumes and it will equal the same value obtained from either of the two linear programming models.

## 5.2. A Water Quality Management Problem

Some linear programming modelling and solution techniques can be demonstrated using the simple water quality management example shown in Figure 4.20. In addition, this example can serve to illustrate how models can help identify just what data are needed and how accurate they must be for the decisions that are being considered.

The stream shown in Figure 4.20 receives wastewater effluent from two point sources located at Sites 1 and 2. Without some wastewater treatment at these sites, the concentration of some pollutant, $P_j$ mg/l, at sites $j = 2$ and 3, will continue to exceed the maximum desired concentration $P_j^{\max}$. The problem is to find the level of wastewater treatment (waste removed) at sites $i = 1$ and 2 that will achieve the desired concentrations at sites $j = 2$ and 3 at a minimum total cost.

This is the classic water quality management problem that is frequently found in the literature, although least-cost objectives have not been applied in practice. There are valid reasons for this that we will review later. Nevertheless, this particular problem can serve to illustrate the development of some linear models for determining data needs, estimating the values of model parameters, and for finding, in this case, cost-effective treatment efficiencies. This problem can also serve to illustrate graphically the general mathematical procedures used for solving linear programming problems.

The first step is to develop a model that predicts the pollutant concentrations in the stream as a function of the pollutants discharged into it. To do this we need some notation. Define $P_j$ to be the pollutant concentration in the stream at site $j$. The total mass per unit time of the pollutant (M/T) in the stream at site $j$ will be its concentration $P_j$ (M/L$^3$) times the streamflow $Q_j$ (L$^3$/T). For example if the concentration is in units of mg/l and

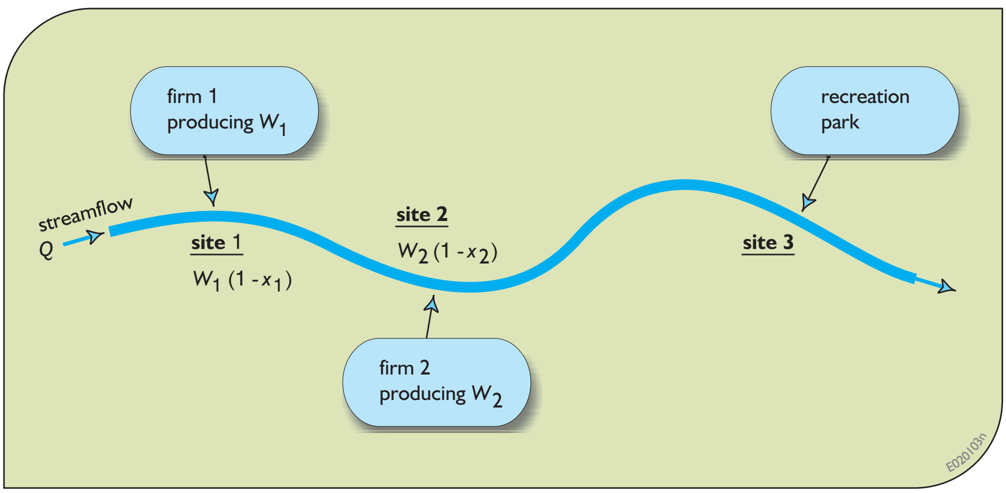



**Figure 4.20.** A stream pollution problem that requires finding the waste removal efficiencies ($x_1$, $x_2$) of wastewater treatment at Sites 1 and 2 that meet the stream quality standards at Sites 2 and 3 at minimum total cost. $W_1$ and $W_2$ are the amounts of pollutant prior to treatment at Sites 1 and 2.

the flow is in terms of $m^3/s$, and mass is to be expressed as kg/day:

$$\text{Mass at site } j \text{ (kg/day)} = P_j \text{(mg/l)}\, Q_j (\text{m}^3/\text{s})\ (10^3 \text{ litres/m}^3)\ (\text{kg}/10^6 \text{ mg})\ (86400 \text{ sec./day}) = 86.4\ P_j Q_j \quad (4.137)$$

Each unit of a degradable pollutant mass in the stream at Site 1 in this example will decrease as it travels downstream to Site 2. Similarly each unit of the pollutant mass in the stream at Site 2 will decrease as it travels downstream to Site 3. The fraction $a_{12}$ of the mass at Site 1 that reaches Site 2 is often assumed to be:

$$a_{12} = \exp(-\mathrm{k}t_{12}) \quad (4.138)$$

where k is a rate constant (1/time unit) that depends on the pollutant and the temperature, and $t_{12}$ is the time (number of time units) it takes a particle of pollutant to flow from Site 1 to Site 2. A similar expression, $a_{23}$, applies for the fraction of pollutant mass at Site 2 that reaches Site 3. The actual concentration at the downstream end of a reach will depend on the streamflow at that site as well as on the initial pollutant mass, the time of travel and rate constant k.

In this example problem, the fraction of pollutant mass at Site 1 that reaches Site 3 is the product of the transfer coefficients $a_{12}$ and $a_{23}$:

$$a_{13} = a_{12}a_{23} \quad (4.139)$$

In general, for any site $k$ between sites $i$ and $j$:

$$a_{ij} = a_{ik}a_{kj} \quad (4.140)$$

Knowing the $a_{ij}$ values for any pollutant and the time of flow $t_{ij}$ permits the determination of the rate constant k for that pollutant and reach, or contiguous series of reaches, from sites $i$ to $j$, using Equation 4.138. If the value of k is 0, the pollutant is called a conservative pollutant; salt is an example of this. Only increased dilution by less polluted water will reduce its concentration.

For the purposes of determining wastewater treatment efficiencies or other capital investments in infrastructure designed to control the pollutant concentrations in the stream, some 'design' streamflow conditions have to be established. Usually the design streamflow condition is set at some very low flow value (e.g., the lowest monthly average flow expected once in twenty years, or the minimum seven-day average flow expected once in ten years). Low design flows are based on the assumption that pollutant concentrations will be higher in low-flow conditions than in higher-flow conditions because of less dilution. While low-flow conditions may not provide as much dilution, they do result in longer travel times, and hence greater reductions in pollutant masses between water quality monitoring sites. Hence the pollutant concentrations may well be greater at some downstream site when the flow conditions are higher than those of the low-flow design value.

In any event, given particular design streamflow and temperature conditions, our first job is to determine the values of these dimensionless transfer coefficients $a_{ij}$. They will be independent of the amount of waste discharged into the stream. To determine both $a_{12}$ and $a_{23}$ in this example problem (Figure 4.20) requires a number of pollutant concentration measurements at Sites 1, 2 and 3 during design streamflow conditions. These measurements of pollutant concentrations must be made just downstream of the wastewater effluent discharge at Site 1, just upstream and downstream of the wastewater effluent discharge at Site 2, and at Site 3.

Assuming no change in streamflow and no extra pollutant entering the reach that begins just downstream of Site 1 and ends just upstream of Site 2, the pollutant concentration $P_2$ just upstream of Site 2 will equal the concentration just downstream of Site 1, $P_1$, times the transfer coefficient $a_{12}$:

$$P_2 = P_1 a_{12} \quad (4.141)$$

Under similar equal flow conditions in the following reach beginning just downstream from Site 2 and extending to Site 3, the pollutant concentration $P_3$ will equal the concentration just downstream of Site 2, $P_2^+$, times the transfer coefficient $a_{23}$.

If the streamflows $Q_i$ and $Q_j$ at sites $i$ and $j$ differ, then the downstream pollutant concentration $P_j$ resulting from an upstream concentration of $P_i$ will be:

$$P_j = P_i a_{ij}(Q_i/Q_j) \quad (4.142)$$

### 5.2.1. Model Calibration

Sample measurements are needed to estimate the values of each reach's pollutant transport coefficients $a_{ij}$. Assume five pairs of sample pollutant concentration measurements have been taken in the two stream reaches during design flow conditions. For this example, also assume that the design streamflow just downstream of Site 1 and just upstream of Site 2 are the same and equal to 12 $m^3/s$. The concentration samples taken just downstream from Site 1 and just upstream of Site 2 during this design flow condition can be used to solve Equation 4.142 after adding error terms. More than one sample is needed to allow for measurement errors and other random effects such as those from wind, incomplete mixing or varying wasteload discharges within a day.

Denote the concentrations of each pair of sample measurements $s$ in the first reach (just downstream of Site 1 and just upstream of Site 2) as $SP_{1s}$ and $SP_{2s}$ and their combined error as $E_s$. Equation 4.142 becomes

$$SP_{2s} + E_s = SP_{1s}\, a_{12}\, (Q_1/Q_2) \tag{4.143}$$

The problem is to find the best estimates of the unknown $a_{12}$. One way to do this is to define 'best' as those values of $a_{12}$ and all $E_s$ that minimize the sum of the absolute values of all the error terms $E_s$. This objective could be written

$$\text{Minimize} \sum_s |E_s| \tag{4.144}$$

The set of Equations 4.143 and 4.144 is an optimization model. If the absolute value signs can be removed, these equations together with constraints that require all unknown variables to be non-negative would form a linear programming model.

The absolute value signs in Equation 4.144 can be removed by writing each error term as the difference between two non-negative variables, $PE_s - NE_s$. Thus for each sample pair $s$:

$$E_s = PE_s - NE_s \tag{4.145}$$

If any $E_s$ is negative, $PE_s$ will be 0 and $-NE_s$ will equal $E_s$. The actual value of $NE_s$ is non-negative. If $E_s$ is positive, it will equal $PE_s$, and $NE_s$ will be 0. The objective function, Equation 4.154, that minimizes the sum of absolute value of error terms, can now be written as one that minimizes the sum of the positive and negative components of $E_s$:

$$\text{Minimize} \sum_s (PE_s + NE_s) \tag{4.146}$$

Equations 4.143 and 4.145, together with objective function 4.146 and a set of measurements, $SP_{1s}$ and $SP_{2s}$, upstream and downstream of the reach between Sites 1 and 2 define a linear programming model that can be solved to find the transfer coefficient $a_{12}$. An example illustrating the solution of this model for the stream reach between Site 1 and just upstream of Site 2 is presented in Box 4.2. Again, the program LINGO (www.lindo.com) is used to solve the models.

Box 4.3 contains the model and solution for the reach beginning just downstream of Site 2 to Site 3. In this reach the design streamflow is 12.5 $m^3/s$ due to the addition of wastewater flow at Site 2.

In this example, based on the solutions for $a_{ij}$ and flows given in Boxes 4.2 and 4.3, $a_{12} = 0.25$, $a_{23} = 0.60$, and thus from Equation 4.140, $a_{12}\, a_{23} = a_{13} = 0.15$.

### 5.2.2. Management Model

Now that these parameter values $a_{ij}$ are known, a water quality management model can be developed. The water quality management problem, illustrated in Figure 4.20, involves finding the fractions $x_i$ of waste removal at sites $i = 1$ and 2 that meet the stream quality standards at the downstream Sites 2 and 3 at a minimum total cost.

The pollutant concentration, $P_2$, just upstream of Site 2 that results from the pollutant concentration at Site 1 equals the total mass of pollutant at Site 1 times the fraction of that mass remaining at Site 2, $a_{12}$, divided by the streamflow just upstream of Site 2, $Q_2$. The total mass of pollutant at Site 1 at the wastewater discharge point is the sum of the mass just upstream of the discharge site, $P_1Q_1$, plus the mass discharged into the stream, $W_1(1 - x_1)$, at Site 1. The parameter $W_1$ is the total amount of pollutant entering the treatment plant at Site 1. Similarly for Site 2. Hence the concentration of pollutant just upstream of Site 2 is:

$$P_2 = [P_1Q_1 + W_1(1-x_1)]a_{12}/Q_2 \tag{4.147}$$

The terms $P_1$ and $Q_1$ are the pollutant concentration ($M/L^3$) and streamflow ($L^3/T$) just upstream of the wastewater discharge outfall at Site 1. Their product is the mass of pollutant at that site per unit time period (M/T).

The pollutant concentration, $P_3$, at Site 3 that results from the pollutant concentration at Site 2 equals the total mass of pollutant at Site 2 times the fraction $a_{23}$. The total mass of pollutant at Site 2 at the wastewater discharge point is the sum of what is just upstream of the discharge site, $P_2Q_2$, plus what is discharged into the stream, $W_2(1-x_2)$. Hence the concentration of pollutant at Site 3 is:

$$P_3 = [P_2Q_2 + W_2(1-x_2)]a_{23}/Q_3 \tag{4.148}$$

**Box 4.2. Calibration of water quality model transfer coefficient parameter $a_{12}$**

E020903b

```
! Calibration of Water Quality Model parameter a12.

Define variables:
SP1(k) = sample pollutant concentration just downstream of site 1 (mg/l).
SP2(k) = sample pollutant concentration just upstream of site 2 (mg/l).
PE(k) = positive error in pollutant conc. sample just upstream of site 2 (mg.l).
NE(k) = negative error in pollutant conc. sample just upstream of site 2 (mg.l).
Qi = streamflow at site i  (i=1, 2), (m3/s).
a12 = pollutant transfer coefficient for stream reach between sites 1 and 2. ;
Sets:
Sample / 1..5 / : PE, NE, SP1, SP2  ;
Endsets
!
Objective: Minimize total sum of positive and negative errors.
;
Min = @sum( Sample:  PE + NE  )
;
!  Subject to constraint for each sample k:
;
@For (Sample:  a12 * SP1  = ( SP2 + PE - NE )* (Q2/Q1));

Data:
SP1 = 232, 256, 220, 192, 204;
SP2 =   55,   67,   53,   50,   51;
Q1 = 12;  !Flow downstream of site 1;  Q2 = 12;   !Flow upstream of site 2;
Enddata


Solution:  a12 = 0.25; Total sum of absolute values of deviations = 10.0
```

Equations 4.147 and 4.148 will become the predictive portion of the water quality management model. The remaining parts of the model include the restrictions on the pollutant concentrations at Sites 2 and 3, and limits on the range of values that each waste removal efficiency $x_i$ can assume.

$$P_j \leq P_j^{\max} \quad \text{for } j = 2 \text{ and } 3 \tag{4.149}$$

$$0 \leq x_i \leq 1.0 \quad \text{for } i = 1 \text{ and } 2. \tag{4.150}$$

Finally, the objective is to minimize the total cost of meeting the stream quality standards $P_2^{\max}$ and $P_3^{\max}$ specified in Equations 4.149. Letting $C_i(x_i)$ represent the wastewater treatment cost function for site $i$, the objective can be written:

$$\text{Minimize } C_1(x_1) + C_2(x_2) \tag{4.151}$$

The complete optimization model consists of Equations 4.147 through 4.151. There are four unknown decision variables, $x_1$, $x_2$, $P_2$, and $P_3$. All variables are assumed to be non-negative.

Some of the constraints of this optimization model can be combined to remove the two unknown concentration values, $P_2$ and $P_3$. Combining Equations 4.147 and 4.149, the concentration just upstream of Site 2 must be no greater than $P_2^{\max}$:

$$[P_1Q_1 + W_1(1-x_1)]a_{12}/Q_2 \leq P_2^{\max} \tag{4.152}$$

Combining Equations 4.148 and 4.149, and using the fraction $a_{13}$ (see Equation 4.139) to predict the contribution of the pollutant concentration at Site 1 on the pollutant concentration at Site 3:

$$\{[P_1Q_1 + W_1(1-x_1)]a_{13} + [W_2(1-x_2)]a_{23}\}/Q_3 \leq P_3^{\max} \tag{4.153}$$

**Box 4.3. Calibration of water quality model transfer coefficient parameter $a_{23}$**

E020903c

```
! Calibration of Water Quality Model parameter a23.

Define variables:
SP2(k) = sample pollutant concentration just downstream of site 2 (mg/l).
SP3(k) = sample pollutant concentration at site 3. (mg/l).
PE(k) = positive error in pollutant conc. sample at site 3  (mg/l).
NE(k) = negative error in pollutant conc. sample at site 3   (mg/l).
Qi = streamflow at site i  (i=2, 3), (m3/s).
a23 = pollutant transfer coefficient for stream reach between sites 2 and 3.
;
Sets:
Sample / 1..5 / : PE, NE, SP2, SP3   ;
Endsets
!
Objective:   Minimize total sum of positive and negative errors.
;
Min = @sum( Sample:  PE + NE  )
;
!  Subject to constraint for each sample k:
;
@For (Sample:  a23 * SP2  = (SP3 + PE – NE )* (Q3/Q2) );

Data:
SP2 = 158,  180,  140,  150,  135;
SP3 =   96,  107,   82,   92,   81;
Q2 = 13;  !Flow just downstream of site 2; Q3 = 13; !Flow at site 3;
Enddata


Solution: a23 = 0.60; Total sum of absolute values of deviations = 6.2
```

Equation 4.153 assumes that each pollutant discharged into the stream can be tracked downstream, independent of the other pollutants in the stream. Alternatively, Equation 4.148 computes the sum of all the pollutants found at Site 2 and then uses that total mass to compute the concentration at Site 3. Both modelling approaches give the same results if the parameter values and cost functions are the same.

To illustrate the solution of either of these models, assume the values of the parameters are as listed in Table 4.22. Rewriting the water quality management model defined by Equations 4.150 to 4.153 and substituting the parameter values in place of the parameters, and recalling that kg/day = 86.4 (mg/l)(m$^3$/s):

$$\text{Minimize } C_1(x_1) + C_2(x_2) \tag{4.154}$$

Subject to:

Water quality constraint at Site 2:

$$[P_1Q_1 + W_1(1-x_1)]a_{12}/Q_2 \le P_2^{\max} \tag{4.155}$$

$$[(32)(10) + 250000(1-x_1)/86.4]0.25/12 \le 20$$

that when simplified is: $x_1 \ge 0.78$

Water quality constraint at Site 3:

$$[P_1Q_1 + W_1(1-x_1)]a_{13} + [W_2(1-x_2)]a_{23}\}/Q_3 \le P_3^{\max} \tag{4.156}$$

$$\{[(32)(10) + 250000(1-x_1)/86.4]0.15 + [80000(1-x_2)/86.4]0.60\}/13 \le 20$$

that when simplified is: $x_1 + 1.28x_2 \ge 1.79$

Restrictions on fractions of waste removal:

$$0 \le x_i \le 1.0 \quad \text{for sites } i = 1 \text{ and } 2 \tag{4.157}$$

**Table 4.22.** Parameter values selected for the water quality management problem illustrated in Figure 4.20.

| | parameter | unit | value | remark |
|---|---|---|---|---|
| flow | $Q_1$ | $m^3/s$ | 10 | flow just upstream of site 1 |
| | $Q_2$ | $m^3/s$ | 12 | flow just upstream of site 2 |
| | $Q_3$ | $m^3/s$ | 13 | flow at park |
| waste | $W_1$ | kg/day | 250,000 | pollutant mass produced at site 1 |
| | $W_2$ | kg/day | 80,000 | pollutant mass produced at site 2 |
| pollutant conc. | $P_1$ | mg/l | 32 | concentration just upstream of site 1 |
| | $P_2$ | mg/l | 20 | maximum allowable concentration upstream of 2 |
| | $P_3$ | mg/l | 20 | maximum allowable concentration at site 3 |
| decay fraction | $a_{12}$ | -- | 0.25 | fraction of site 1 pollutant mass at site 2 |
| | $a_{13}$ | -- | 0.15 | fraction of site 1 pollutant mass at site 3 |
| | $a_{23}$ | -- | 0.60 | fraction of site 2 pollutant mass at site 2 |

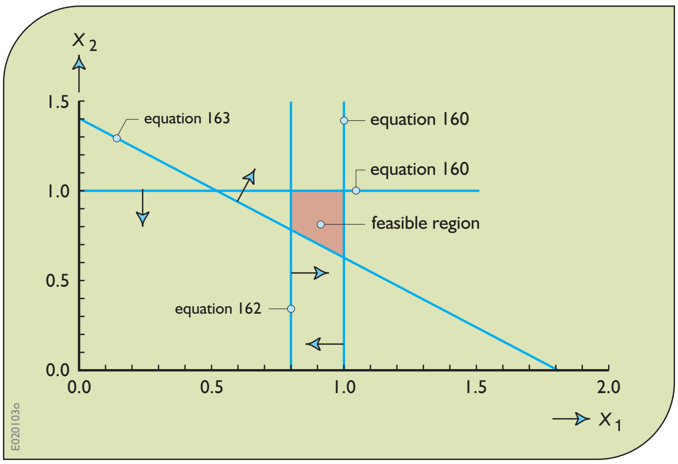



**Figure 4.21.** Plot of the constraints of water quality management model identifying those values of the unknown (decision) variables $x_1$ and $x_2$ that satisfy all the constraints.

The feasible combinations of $x_1$ and $x_2$ can be shown on a graph, as in Figure 4.21. This graph is a plot of each constraint, showing the boundaries of the region of combinations of $x_1$ and $x_2$ that satisfy all the constraints. This red shaded region is called the feasible region.

To find the least-cost solution we need the cost functions $C_1(x_1)$ and $C_2(x_2)$ in Equations 4.151 or 4.154. Suppose these functions are not known. Can we determine the least-cost solution without knowing these costs? Models like the one just developed can be used to determine just how accurate these cost functions (or any model parameters) need to be for the decisions being considered.

While the actual cost functions are not known, their general form can be assumed, as shown in Figure 4.22. Since the wasteloads produced at Site 1 are substantially greater than those produced at Site 2, and given similar site, transport, labour, and material cost conditions, it seems reasonable to assume that the cost of providing a specified level of treatment at Site 1 would exceed (or certainly be no less than) the cost of providing the same specified level of treatment at Site 2. It would also seem the marginal cost at Site 1 would be greater than, or at least no less than, the marginal cost at Site 2 for the same amount of treatment. The relative positions of the cost functions shown in Figure 4.22 are based on these assumptions.

Rewriting the cost function, Equation 4.154, as a linear function converts the model defined by Equations 4.150 through 4.151 into a linear programming model. For this example problem, the linear programming model can be written as:

$$\text{Minimize } c_1x_1 + c_2x_2 \qquad (4.158)$$

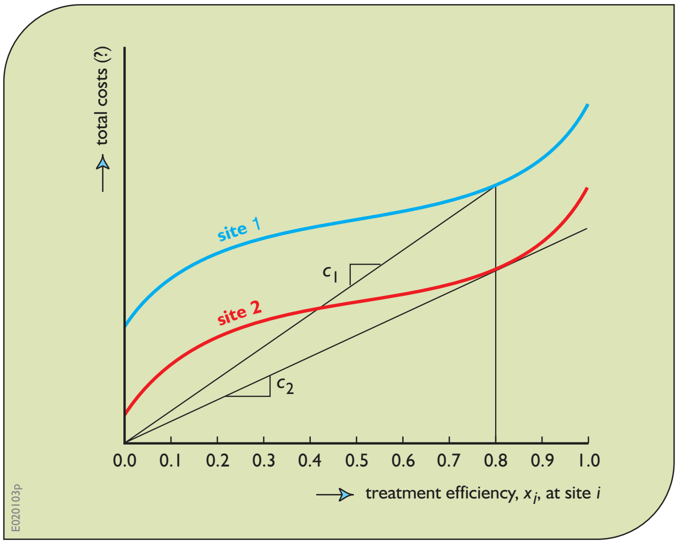



**Figure 4.22.** General form of total cost functions for wastewater treatment efficiencies at Sites 1 and 2 in Figure 4.20. The dashed straight-line slopes $c_1$ and $c_2$ are the average cost per unit (%) removal for 80% treatment. The actual average costs clearly depend on the values of the waste removal efficiencies $x_1$ and $x_2$ respectively.

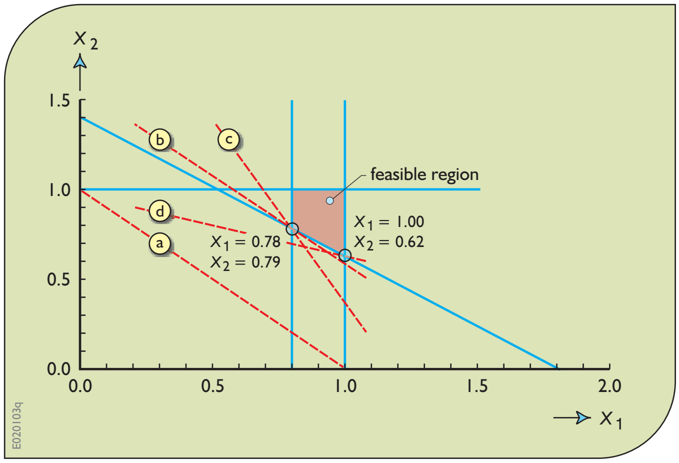



**Figure 4.23.** Plots of various objective functions (dashed lines) together with the constraints of the water quality management model.

Subject to:

$$x_1 \geq 0.78 \tag{4.159}$$

$$x_1 + 1.28\, x_2 \geq 1.79 \tag{4.160}$$

$$0 \leq x_i \leq 1.0 \quad \text{for } i = 1 \text{ and } 2 \tag{4.161}$$

where the values of $c_1$ and $c_2$ depend on the values of $x_1$ and $x_2$ and both pairs are unknown. Even if we knew the values of $x_1$ and $x_2$ before solving the problem, in this example the cost functions themselves (Figure 4.22) are unknown. Hence, we cannot determine the values of the marginal costs $c_1$ and $c_2$. However, we might be able to judge which will likely be greater than the other for any particular values of the decision-variables $x_1$ and $x_2$.

First, assume $c_1$ equals $c_2$. Let $c_1x_1 + c_2x_2$ equal $c$ and assume $c/c_1 = 1$. Thus $x_1 + x_2 = 1.0$. This line can be plotted onto the graph in Figure 4.23, as shown by line 'a' in Figure 4.23.

Line 'a' in Figure 4.23 represents equal values for $c_1$ and $c_2$, and the total cost, $c_1x_1 + c_2x_2$, equal to 1. Keeping the slope of this line constant and moving it upward, representing increasing total costs, to line 'b', where it covers the nearest point in the feasible region, will identify the least-cost combination of $x_1$ and $x_2$, again assuming the marginal costs are equal. In this case the solution is approximately 80% treatment at both sites.

Note this particular least-cost solution applies for any value of $c_1$ greater than $c_2$ (for example line 'c' in Figure 4.23). If the marginal cost of 80% treatment at Site 1 is no less than the marginal cost of 80% treatment at Site 2, then $c_1 \geq c_2$ and indeed the 80% treatment efficiencies will meet the stream standards for the design streamflow and wasteload conditions at a total minimum cost. In fact, from Figure 4.23 and Equation 4.160, it is clear that $c_2$ has to exceed $c_1$ by a multiple of 1.28 before the least-cost solution changes to another solution. For any other assumption regarding $c_1$ and $c_2$, 80% treatment at both sites will result in a least-cost solution to meeting the water quality standards for those design wasteload and streamflow conditions.

If $c_2$ exceeds $1.28c_1$, as illustrated by line 'd', then the least-cost solution would be $x_1 = 100\%$ and $x_2 = 62\%$. Clearly, in this example the marginal cost, $c_1$, of providing 100% wasteload removal at Site 1 will exceed the marginal cost, $c_2$, of 60% removal at Site 2, and hence, that combination of efficiencies would not be a least-cost one. Thus we can be confident that the least-cost solution is to remove 80% of the waste produced at both waste-generating sites.

Note the least-cost wasteload removal efficiencies have been determined without knowing the cost functions. Why spend money defining these functions more precisely? The answer: costs need to be known for financial planning, if not for economic analyses. No doubt the actual costs of installing the least-cost treatment efficiencies of 80% will have to be determined for issuing bonds or making other arrangements for paying the costs. However, knowing the least-cost removal efficiencies means we do not have to spend money defining the entire cost functions $C_i(x_i)$. Estimating the construction and operating costs of achieving just one wastewater removal efficiency at each site, namely 80%, should be less expensive than defining the total costs for a range of practical treatment plant efficiencies that would be required to define the total cost functions, such as shown in Figure 4.22.

Admittedly this example is relatively simple. It will not always be possible to determine the 'optimal' solutions to linear programming problems, or other optimization problems, without knowing more about the objective function than was assumed for this example. However, this exercise illustrates the use of modelling for purposes other than finding good or 'optimal' solutions. Models can help define the necessary precision of the data needed to find those good solutions.

Modelling and data collection and analysis should take place simultaneously. All too often planning exercises are divided into two stages: data collection and then analysis. Until one knows what data one will need, and how accurate those data must be, one should not spend money and time collecting them. Conversely, model development in the absence of any knowledge of the availability and cost of obtaining data can lead to data requirements that are costly, or even impossible, to obtain, at least in the time available for decision-making. Data collection and model development are activities that should be performed simultaneously.

Because software is widely available to solve linear programming programs, because these software programs can solve very large problems containing thousands of variables and constraints, and finally because there is less chance of obtaining a local 'non-optimal' solution when the problem is linear (at least in theory), there is an incentive to use linear programming to solve large optimization problems. Especially for large optimization problems, linear programming is often the only practical alternative. Yet models representing particular water resources systems may not be linear. This motivates the use of methods that can approximate non-linear functions with linear ones.

The following simple groundwater supply problem illustrates the application of some linearization methods commonly applied to non-linear separable functions – functions of only one unknown variable.

These methods typically increase the number of variables and constraints in a model. Some of these methods require integer variables, or variables that can have values of only 0 or 1. There is a practical limit on the number of integer variables any linear programming software program can handle. Hence, for large models there may be a need to perform some preliminary screening designed to reduce the number of alternatives that should be considered in more detail. This example can be used to illustrate an approach to preliminary screening.

## 5.3. A Groundwater Supply Example

Consider a water-using industry that plans to obtain water from a groundwater aquifer. Two wellfield sites have been identified. The first question is how much the water will cost, and the second, given any specified amount of water delivered to the user, is how much should come from each wellfield. This situation is illustrated in Figure 4.24.

Wells and pumps must be installed and operated to obtain water from these two wellfields. The annual cost of wellfield development will depend on the pumping capacity of the wellfield. Assume that the annual costs associated with various capacities $Q_A$ and $Q_B$ for Wellfields

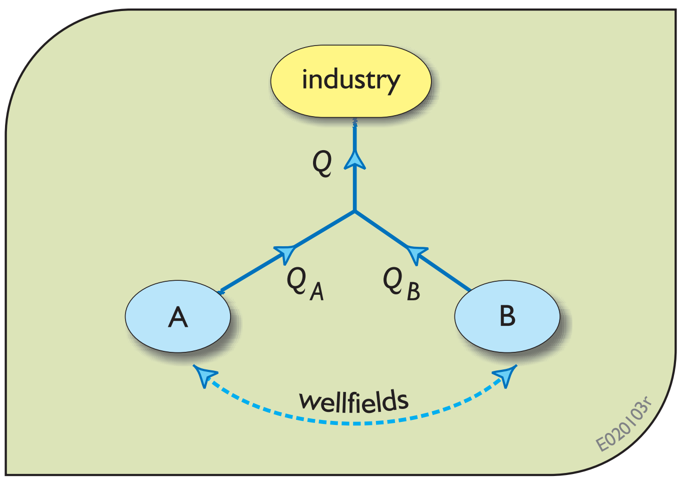



**Figure 4.24.** Schematic of a potential groundwater supply system that can serve a water using industry. The unknown variables are the flows, $Q_A$ and $Q_B$, from each wellfield.

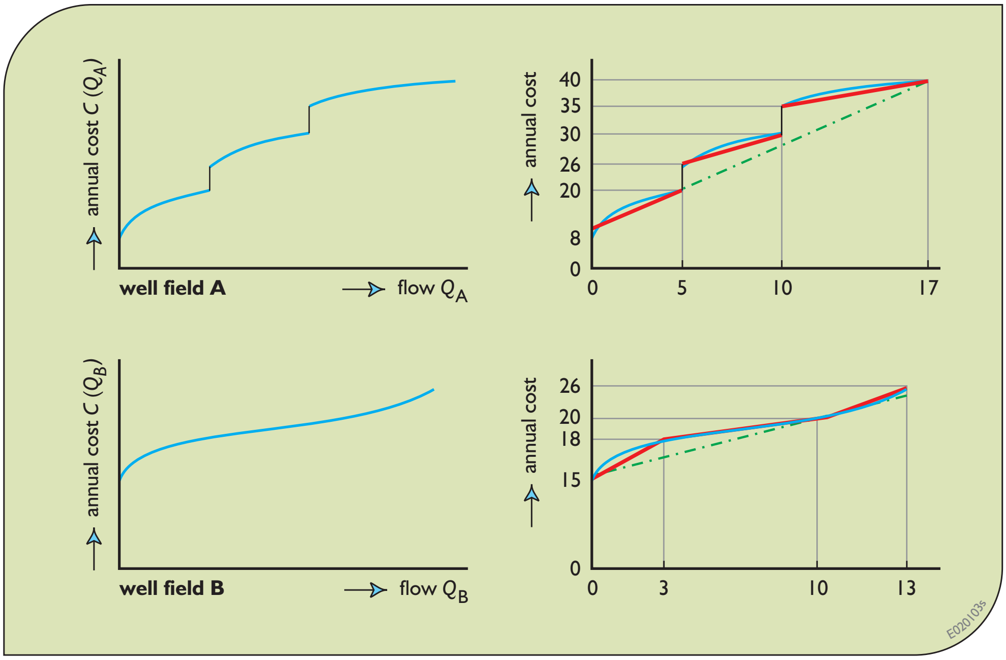



**Figure 4.25.** Annual cost functions associated with the Wellfields A and B as shown in Figure 4.24. The actual functions are shown on the left, and two sets of piecewise linear approximations are shown on the right.

A and B respectively are as shown in Figure 4.25. These are non-linear functions that contain both fixed and variable costs and hence are discontinuous. The fixed costs result from the fact that some of the components required for wellfield development come in discrete sizes. As indicated in the figure, the maximum flow capacity of Wellfields A and B are 17 and 13, respectively.

In Figure 4.25, the non-linear functions on the left have been approximated by piecewise linear functions on the right. This is a first step in linearizing non-linear separable functions. Increasing the number of linear segments can reduce the difference between the piecewise linear approximation of the actual non-linear function and the function itself. At the same time it will increase the number of variables and possibly constraints.

When approximating a non-linear function by a series of straight lines, model developers should consider two factors. The first is just how accurate need be the approximation of the actual function for the decisions that will be made, and second is just how accurate is the actual (in this case non-linear) function in the first place. There is little value in trying to eliminate relatively small errors caused by the linearization of a function when the function itself is highly uncertain. Most cost and benefit functions, especially those associated with future activities, are indeed uncertain.

### 5.3.1. A Simplified Model

Two sets of approximations are shown in Figure 4.25. Consider first the approximations represented by the light blue dot–dash lines. These are very crude approximations – a single straight line for each function. In this example these straight-line cost functions are lower bounds of the actual non-linear costs. Hence, the actual costs may be somewhat higher than those identified in the solution of a model.

Using the blue dot–dash linear approximations in Figure 4.25, the linear programming model can be written as follows:

$$\text{Minimize CostA} + \text{CostB} \tag{4.162}$$

Subject to:

$$\text{CostA} = 8I_A + [(40-8)/17]Q_A \quad \{\text{linear approximation of } C(Q_A)\} \tag{4.163}$$

$$\text{CostB} = 15I_B + [(26-15)/13]Q_B \quad \{\text{linear approximation of } C(Q_B)\} \tag{4.164}$$

$$I_A,\ I_B \text{ are 0, 1 integer variables} \tag{4.165}$$

$$Q_A \leq 17I_A \ \{\text{limits } Q_A \text{ to 17 and forces } I_A = 1 \text{ if } Q_A > 0\} \tag{4.166}$$

$$Q_B \leq 13I_B \ \{\text{limits } Q_B \text{ to 13 and forces } I_B = 1 \text{ if } Q_B > 0\} \tag{4.167}$$

$$Q = Q_A + Q_B \text{ \{mass balance\}} \quad (4.168)$$

$$Q, Q_A, Q_B \geq 0 \text{ \{non-negativity of all decision variables\}} \quad (4.169)$$

$$Q = \text{some specified amount from 0 to 30.} \quad (4.170)$$

The expressions within the square brackets, [ ], above represent the slopes of the dot–dash linear approximations of the cost functions. The integer 0, 1 variables are required to include the fixed costs in the model.

Solving this model for various values of the water demand $Q$ provides some interesting results. Again, they are based on the dot–dash linear cost functions in Figure 4.25. As $Q$ increases from 0 to just under 6.8, all the water will come from the less expensive Wellfield A. For any $Q$ from 6.8 to 13, Wellfield B becomes less expensive and all the water will come from it. For any $Q$ greater than the capacity of Wellfield B of 13 but no greater than the capacity of Wellfield A, 17, all of it will come from Wellfield A. Because of the fixed costs, it is cheaper to use one rather than both wellfields. Beyond $Q = 17$, the maximum capacity of A, water needs to come from both wellfields. Wellfield B will pump at its capacity, 13, and the additional water will come from Wellfield A.

Figure 4.26 illustrates these solutions. One can understand why in situations of increasing demands for $Q$ over time, capacity-expansion modelling might be useful. One would not close down a wellfield once developed, just to achieve what would have been a least-cost solution if the existing wellfield had not been developed.

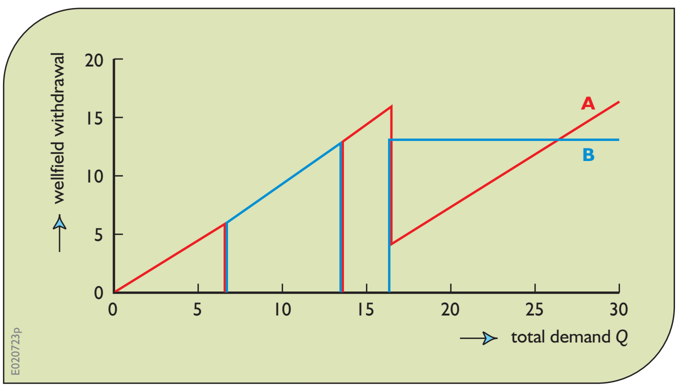



**Figure 4.26.** Least-cost wellfield use given total demand $Q$ based on model defined by Equations 4.162 to 4.170.

### 5.3.2. A More Detailed Model

A more accurate representation of these cost functions may change these solutions for various values of $Q$, although not significantly. However consider the more accurate cost minimization model that includes the red solid-line piecewise linearizations shown in Figure 4.25.

$$\text{Minimize CostA} + \text{CostB} \quad (4.171)$$

Subject to:

$$\text{CostA} = \{8I_{A1} + [(20-8)/5]Q_{A1}\} + \{26I_{A2} + [(30-26)/(10-5)]Q_{A2}\} + \{35I_{A3} + [(40-35)/(17-10)]Q_{A3}\} \text{ \{linear approximation of } C(Q_A)\text{\}} \quad (4.172)$$

$$\text{CostB} = \{15I_{B1} + [(18-15)/3]Q_{B1}\} + \{18I_{B2} + [(20-18)/(10-3)]Q_{B2} + [(26-20)/(13-10)]Q_{B3}\} \text{ \{linear approximation of } C(Q_B)\text{\}} \quad (4.173)$$

$$Q_A = Q_{A1} + (5I_{A2} + Q_{A2}) + (10I_{A3} + Q_{A3}) \text{ \{}Q_A\text{ defined\}} \quad (4.174)$$

$$Q_B = Q_{B1} + (3I_{B2} + Q_{B2} + Q_{B3}) \text{ \{}Q_B\text{ defined\}} \quad (4.175)$$

$I_{Ai}$, $I_{Bi}$ are 0, 1 integer variables for all segments $i$ (4.176)

$$Q_{A1} \leq 5I_{A1},\ Q_{A2} \leq (10-5)I_{A2},\ Q_{A3} \leq (17-10)I_{A3}$$

{limits $Q_{Ai}$ to width of segment $i$ and forces $I_{Ai} = 1$ if $Q_{Ai} > 0$} (4.177)

$$I_{A1} + I_{A2} + I_{A3} \leq 1 \quad (4.178)$$

{limits solution to at most only one cost function segment $i$}

$$Q_{B1} \leq 3I_{B1},\ Q_{B2} \leq (10-3)I_{B2},\ Q_{B3} \leq (13-10)I_{B2} \quad (4.179)$$

{limits $Q_{Bi}$ to width of segment $i$ and forces $I_{Bi} = 1$ if $Q_{Bi} > 0$}

$$I_{B1} + I_{B2} \leq 1 \quad (4.180)$$

{limits solution to at most only the first segment or to the second and third segments of the cost function. Note that a 0, 1 integer variable for the fixed cost of the third segment of this function is not needed since its slope exceeds that of the second segment. However the flow, $Q_{B3}$, in that segment must be bounded using the integer 0, 1 variable, $I_{B2}$, associated with the second segment, as shown in the third of Equations 4.179}

$$Q = Q_A + Q_B \text{ \{mass balance\}} \tag{4.181}$$

$$Q,\ Q_A,\ Q_B \geq 0 \text{ \{non-negativity of all decision variables\}} \tag{4.182}$$

$$Q = \text{some specified amount from 0 to 30.} \tag{4.183}$$

The solution to this model, shown in Figure 4.27, differs from the solution of the simpler model, but only in the details. Wellfield A supplies all the water for $Q \leq 4.3$. For values of $Q$ between 4.4 and 13 all the water comes from Wellfield B. For values of $Q$ in excess of 13 to 14.8, the capacity of Wellfield B remains at 13 and Wellfield A provides the additional amount of needed capacity over 13. For $Q = 14.9$ to 17, the capacity of Wellfield B drops to 0 and the capacity of Wellfield A increases from 14.9 to 17. For values of $Q$ between 17 and 18 Wellfield B provides 13, its capacity, and the capacity of A increases from 4 to 5. For values of $Q$ from 18.1 to 27, Wellfield B provides 10, and Wellfield A increases from 8.1 to 17.

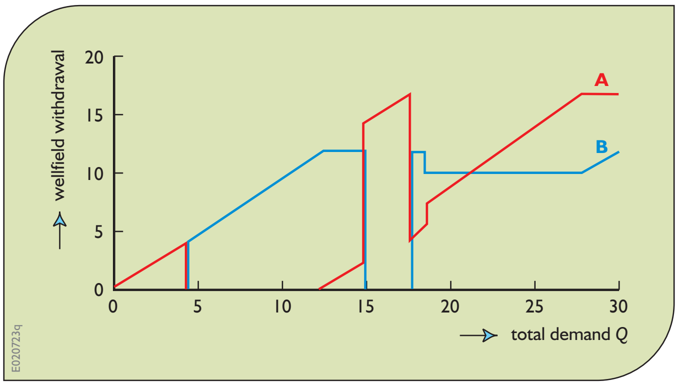



**Figure 4.27. Least-cost wellfield use given total demand *Q* based on Equations 4.171 to 4.183.**

For values of $Q$ in excess of 27, Wellfield A remains at 17, its capacity, and Wellfield B increases from 10 to 13.

As in the previous example, this shows the effect on the least-cost solution when one cost function has relatively lower fixed and higher variable costs compared with another cost function having relatively higher fixed and lower variable costs.

### 5.3.3. An Extended Model

In this example, the simpler model (Equations 4.162 to 4.170) and the more accurate model (Equations 4.171 to 4.183) provided essentially the same allocations of wellfield capacities associated with a specified total capacity $Q$. If the problem contained a larger number of wellfields, the simpler (and smaller) model might have been able to eliminate some of these wellfields from further consideration. This would reduce the size of any new model that approximates the cost functions of the remaining wellfields more accurately.

The model just described, like the capacity-expansion model and water quality management model, is another example of a cost-effective model. The objective was to find the least-cost way of providing a specified amount of water to a water user. Next, consider a cost–benefit analysis in which the question is just how much water the user should use. To address this question we assume the user has identified the annual benefits associated with various amounts of water. The annual benefit function, $B(Q)$, and its piecewise linear approximations, are shown in Figure 4.28.

The straight, blue, dot–dash–dash linear approximation of the benefit function shown in Figure 4.28 is an

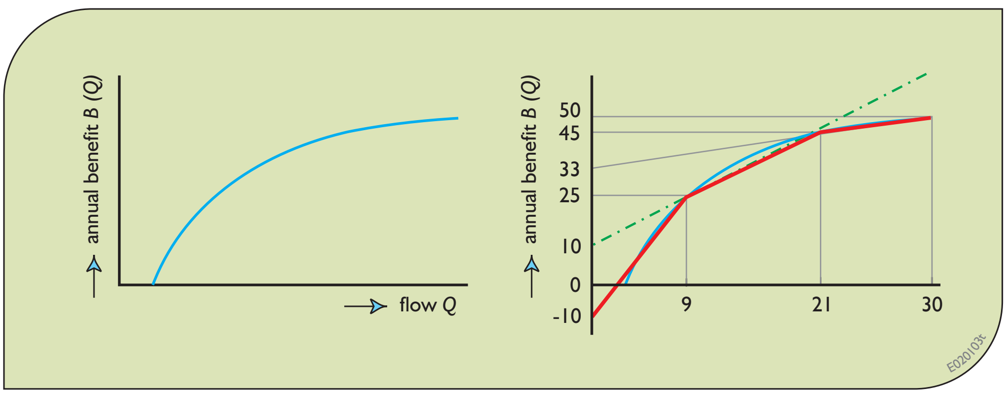



**Figure 4.28. Benefit function of the amount of water provided to the water user. Piecewise linear approximations of that function of flow are shown on the right.**

upper bound of the benefits. Incorporating it into a model that uses the dot–dash linear lower bound approximations of each cost function, as shown in Figure 4.25, will produce an optimistic solution. It is unlikely that the value of $Q$ that is based on more accurate and thus less optimistic benefit and cost functions will be any greater than the one identified by this simple optimistic model. Furthermore, if any wellfield is not in the solution of this optimistic model, with some care we might be able to eliminate that wellfield from further consideration when developing a more accurate model.

Any component of a water resources system that does not appear in the solution of a model that includes optimistic approximations of performance measures that are to be maximized, such as benefits, or that are to be minimized, such as costs, are candidates for omission in any more detailed model. This is an example of the process of preliminary screening.

The model defined by Equations 4.162 through 4.170 can now be modified. Equation 4.170 is eliminated and the cost-minimization objective Equation 4.162 is replaced with:

$$\text{Maximize } Benefits - (\text{CostA} + \text{CostB}) \tag{4.184}$$

where

$$Benefits = 10 + [(45 - 25)/(21 - 9)]Q \quad \{\text{linear approximation of } B(Q)\} \tag{4.185}$$

The solution of this model, Equations 4.163 to 4.169, 4.184, and 4.185 (plus the condition that the fixed benefit of 10 only applies if $Q > 0$, added because it is clear the benefits would be 0 with a $Q$ of 0) indicates that only Wellfield B needs to be developed, and at a capacity of 10. This would suggest that Wellfield A can be omitted in any more detailed modelling exercise. To see if this assumption, in this example, is valid, consider the more detailed model that incorporates the red, solid-line linear approximations of the cost and benefit functions shown in Figures 4.25 and 4.28.

Note that the approximation of the generally concave benefit function in Figure 4.28 will result in negative values of the benefits for small values of $Q$. For example, when the flow $Q$, is 0 the approximated benefits are –10. Yet the actual benefits are 0 as shown in the left part of Figure 4.28. Modelling these initial fixed benefits the same way as the fixed costs have been modelled, using another 0, 1 integer variable, would allow a more accurate representation of the actual benefits for small values of $Q$.

Alternatively, to save having to add another integer variable and constraint to the model, one can allow the benefits to be negative. If the model solution shows negative benefits for some small value of $Q$, then obviously the more preferred value of $Q$, and benefits, would be 0. This more approximate trial-and-error approach is often preferred in practice, especially when a model contains a large number of variables and constraints. This is the approach taken here.

### 5.3.4. Piecewise Linearization Methods

There are a number of ways of modelling the piecewise linear concave benefit function shown on the right side of Figure 4.28. Several are defined in the next several sets of equations. Each method will result in the same model solution.

One approach to modelling the concave benefit function is to define a new unrestricted (possibly negative-valued) variable. Let this variable be *Benefits*. When being maximized this variable cannot exceed any of the linear functions that bound the concave benefit function:

$$Benefits \le -10 + [(25 - (-10))/9]Q \tag{4.186}$$

$$Benefits \le 10 + [(45 - 25)/(21-9)]Q \tag{4.187}$$

$$Benefits \le 33 + [(50 - 45)/(30-21)]Q \tag{4.188}$$

Since most linear programming algorithms assume all unknown variables are non-negative (unless otherwise specified), unrestricted variables, such as *Benefits,* can be replaced by the difference between two non-negative variables, such as *Pben – Nben*. *Pben* will equal *Benefits* if its value is greater than 0. Otherwise –*Nben* will equal *Benefits*. Thus in place of *Benefits* in the Equations 4.186 to 4.188, and those below, one can substitute *Pben – Nben*.

Another modelling approach is to divide the variable $Q$ into parts, $q_i$, one for each segment $i$ of the function. These parts sum to $Q$. Each $q_i$, ranges from 0 to the width of the user-defined segment $i$. Thus for the piecewise linear benefit function shown on the right of Figure 4.28:

$$q_1 \le 9 \tag{4.189}$$

$$q_2 \le 21 - 9 \tag{4.190}$$

$$q_3 \le 30 - 21 \tag{4.191}$$

and

$$Q = q_1 + q_2 + q_3 \tag{4.192}$$

The linearized benefit function can now be written as the sum over all three segments of each segment slope times the variable $q_i$:

$$\begin{aligned} Benefits = -10 &+ [(25 + 10)/9]q_1 + [(45 - 25)/(21 - 9)]q_2 \\ &+ [(50 - 45)/(30 - 21)]q_3 \end{aligned} \tag{4.193}$$

Since the function being maximized is concave (decreasing slopes as $Q$ increases), we are assured that each $q_{i+1}$ will be greater than 0 only if $q_i$ is at its upper limit, as defined by constraint Equations 4.189 to 4.191.

A third method is to define unknown weights $w_i$ associated with the breakpoints of the linearized function. The value of $Q$ can be expressed as the sum of a weighted combination of segment endpoint values. Similarly, the benefits associated with $Q$ can be expressed as a weighted combination of the benefits evaluated at the segment endpoint values. The weights must also sum to 1.

Hence, for this example:

$$Benefits = (-10)w_1 + 25w_2 + 45w_3 + 50w_4 \tag{4.194}$$

$$Q = 0w_1 + 9w_2 + 21w_3 + 30w_4 \tag{4.195}$$

$$1 = w_1 + w_2 + w_3 + w_4 \tag{4.196}$$

For this method to provide the closest approximation of the original non-linear function, the solution must include no more than two non-zero weights and those non-zero weights must be adjacent to each other. Since a concave function is to be maximized, this condition will be met, since any other situation would yield less benefits.

The solution to the more detailed model defined by Equations 4.184, 4.172 to 4.182, and either 4.186 to 4.188, 4.189 to 4.193, or 4.194 to 4.196, indicates a value of 10 for $Q$ will result in the maximum net benefits. This flow is to come from Wellfield B. This more precise solution is identical to the solution of the simpler model. Clearly the simpler model could have successfully served to eliminate Wellfield A from further consideration.

## 5.4. A Review of Linearization Methods

This section presents a review of the piecewise linearization methods just described and some other approaches for incorporating special conditions into linear programming models.

### If–then–else conditions

There exist a number of ways 'if–then–else' and 'and' and 'or' conditions (that is, decision trees) can be included in linear programming models. To illustrate some of them, assume $X$ is an unknown decision-variable in a model whose value may depend on the value of another unknown decision-variable $Y$. Assume the maximum value of $Y$ would not exceed $Y$max and the maximum value of $X$ would not exceed $X$max. These upper bounds and all the linear constraints representing 'if–then–else' conditions must not restrict the values of the original decision-variable $Y$. Four 'if–then–else (with 'and/or') conditions are presented below using additional integer 0.1 variables, denoted by $Z$. All the $X$, $Y$ and $Z$ variables in the constraints below are assumed to be unknown. These constraints would be included in the any linear programming model where the particular 'if–then–else' conditions apply.

These illustrations are not unique. At the boundaries of the 'if' constraints in the examples below, either of the 'then' or 'else' conditions can apply.

a) If $Y \le 50$ then $X \le 10$, else $X \ge 15$.
   Define constraints:
   $Y \le 50Z + Y\text{max}(1 - Z)$ where $Z$ is a 0, 1 integer variable.
   $Y \ge 50(1 - Z)$
   $X \le 10Z + X\text{max}(1 - Z)$
   $X \ge 15(1 - Z)$
b) If $Y \le 50$ then $X \le Y$, else $X \ge Y$.
   Define constraints:
   $Y \ge 50Z$ where $Z$ is a 0, 1 integer variable.
   $Y \le 50(1 - Z) + Y\text{max}Z$
   $X \le Y + X\text{max}Z$
   $X \ge Y - Y\text{max}(1 - Z)$
c) If $Y \le 20$ or $Y \ge 80$ then $X = 5$, else $X \ge 10$.
   Define constraints:
   $Y \le 20Z_1 + 80Z_2 + Y\text{max}(1 - Z_1 - Z_2)$
   $Y \ge 20Z_2 + 80(1 - Z_1 - Z_2)$
   $Z_1 + Z_2 \le 1$ where each $Z$ is a 0, 1 integer variable.
   $X \le 5(Z_1 + (1 - Z_1 - Z_2)) + X\text{max}Z_2$
   $X \ge 5(Z_1 + (1 - Z_1 - Z_2))$
   $X \ge 10Z_2$

d) If $20 \leq Y \leq 50$ or $60 \leq Y \leq 80$, then $X \leq 5$, else $X \geq 10$.

Define constraints:

$Y \leq 20Z_1 + 50Z_2 + 60Z_3 + 80Z_4 + Y\max(1 - Z_1 - Z_2 - Z_3 - Z_4)$

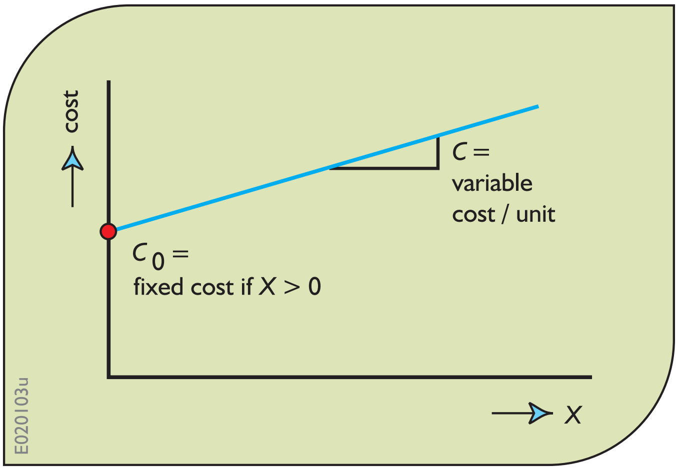
$Y \geq 20Z_2 + 50Z_3 + 60Z_4 + 80(1 - Z_1 - Z_2 - Z_3 - Z_4)$

$Z_1 + Z_2 + Z_3 + Z_4 \leq 1$ where each $Z$ is a 0, 1 integer variable.

$X \leq 5\,(Z_2 + Z_4) + X\max{}^*(1 - Z_2 - Z_4)$

$X \geq 10^*\,((Z_1 + Z_3) + (1 - Z_1 - Z_2 - Z_3 - Z_4))$

### Fixed costs in cost functions:




Cost $= C_0 + CX$ if $X > 0$,
$= 0$ otherwise.

To include these fixed costs in a LP model, define Cost $= C_0I + CX$ and constrain $X \leq MI$ where M is the maximum value of $X$, and $I$ is an unknown 0, 1 variable.

### Minimizing the maximum or maximizing the minimum

Let the set of variables be $\{X_1, X_2, X_3, \ldots, X_n\}$

Minimize maximum $\{X_1, X_2, X_3, \ldots, X_n\}$ is equivalent to:

Minimize $U$ subject to $U \geq X_j$, $j = 1, 2, 3, \ldots, n$.

Maximize minimum $\{X_1, X_2, X_3, \ldots, X_n\}$ is equivalent to:

Maximize $L$ subject to $L \leq X_j$, $j = 1, 2, 3, \ldots, n$.

### Minimizing the absolute value of the difference between two unknown non-negative variables:

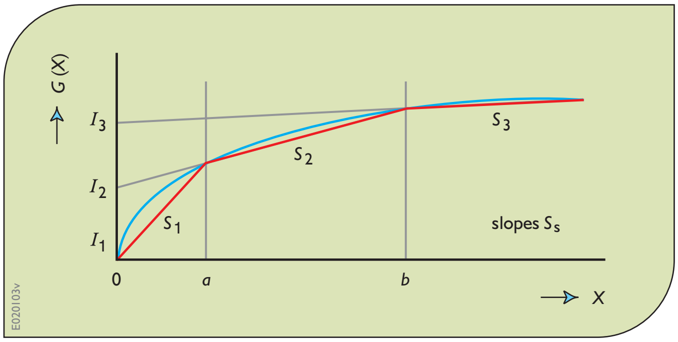
Minimize $|X - Y|$ is equivalent to

Minimize $D$ subject to $X - Y \leq D$; $Y - X \leq D$; $X, Y, D \geq 0$.

or Minimize $(PD+ND)$

subject to: $X - Y = PD - ND$; $PD, ND, X, Y \geq 0$.

### Minimizing convex functions or maximizing concave functions

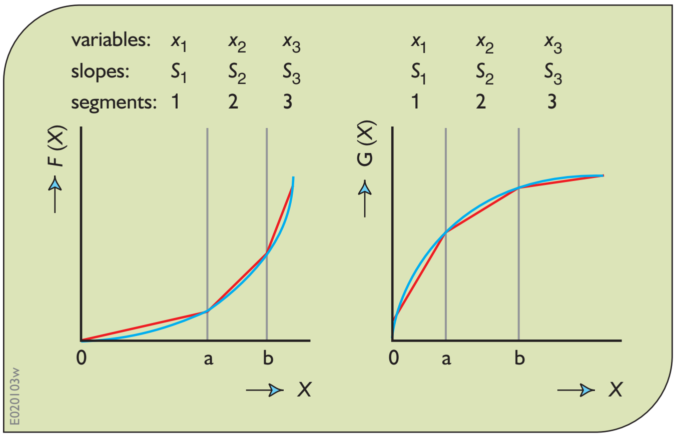



Maximize $G(X) \cong$ Maximize $B$

Subject to: $I_1 + S_1X \geq B$

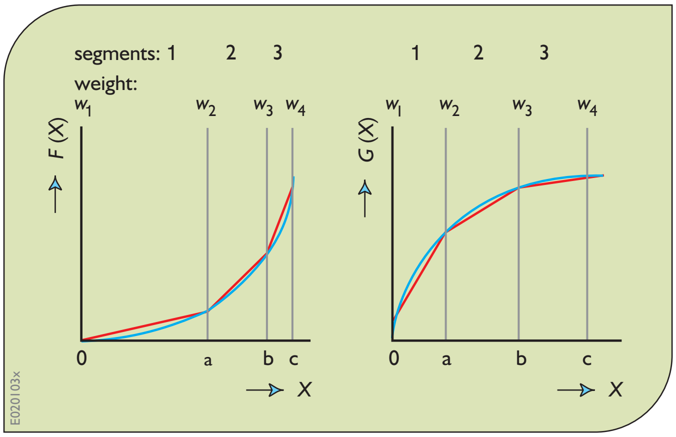
$I_2 + S_2X \geq B$

$I_3 + S_3X \geq B$




$F(X) \cong S_1x_1 + S_2x_2 + S_3x_3$;
$G(X) \cong S_1x_1 + S_2x_2 + S_3x_3$

$X = x_1 + x_2 + x_3$; $x_1 \leq a$; $x_2 \leq b - a$



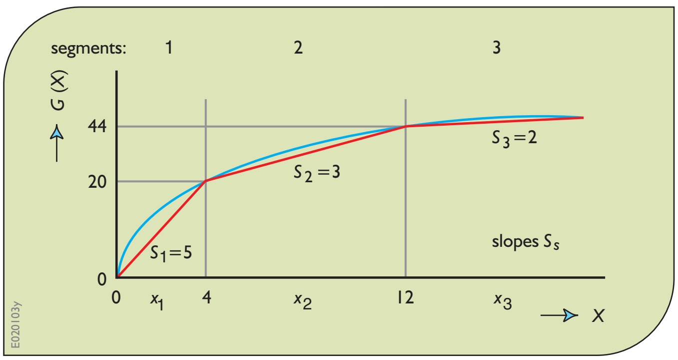

Minimize $F(X) \cong F(0)w_1 + F(a)w_2 + F(b)w_3 + F(c)w_4$

Maximize $G(X) \cong G(0)w_1 + G(a)w_2 + G(b)w_3 + G(c)w_4$

Subject to:

$$X = 0w_1 + aw_2 + bw_3 + cw_4\,;$$
$$w_1 + w_2 + w_3 + w_4 = 1$$

## Minimizing concave functions or maximizing convex functions



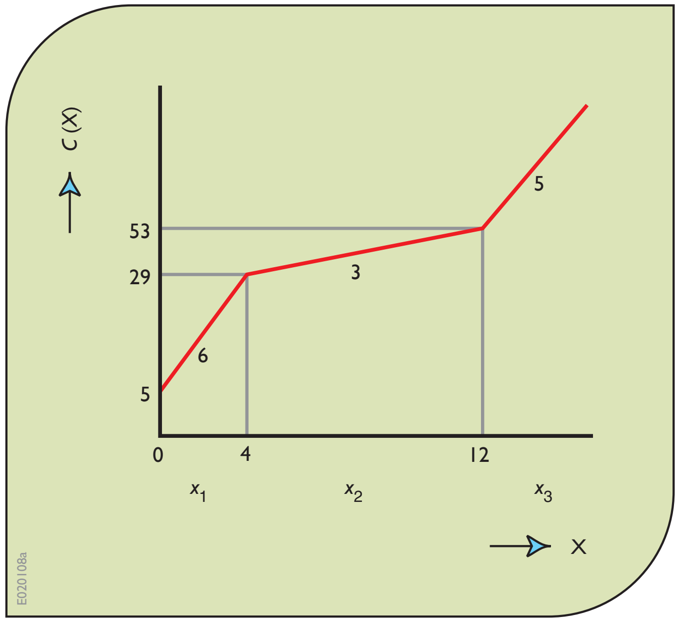

Minimize $G(X) \cong 5x_1 + (20z_2 + 3x_2) + (44z_3 + 2x_3)$

Subject to:

$$x_1 + (4z_2 + x_2) + (12z_3 + x_3) = X;$$
$$z_s = 0 \text{ or } 1\ \forall s;$$

$$x_1 \le 4z_1; \quad x_2 \le 8z_2; \quad x_3 \le 99\, z_3; \quad z_1 + z_2 + z_3 = 1.$$

## Minimizing or maximizing combined concave–convex functions




Maximize $C(X) \cong (5z_1 + 6x_1 + 3x_2) + (53z_3 + 5x_3)$

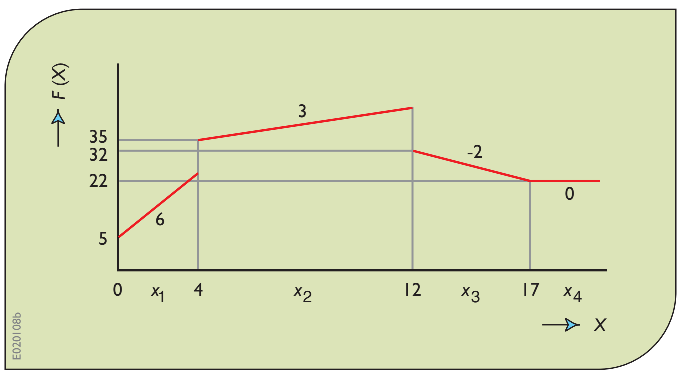
Subject to: $(x_1 + x_2) + (12z_3 + x_3) = X;$

$x_1 \le 4z_1;\quad x_2 \le 8z_1;\quad x_3 \le 99z_3;\quad z_1 + z_3 = 1;$
$z_1, z_3 = 0, 1.$

Minimize $C(X) \cong (5z_1 + 6x_1) + (29z_2 + 3x_2 + 5x_3)$

Subject to: $x_1 + (4z_2 + x_2 + x_3) = X;$

$x_1 \le 4z_1;\quad x_2 \le 8z_2;\quad x_3 \le 99\, z_2;\quad z_1 + z_2 \le 1;$
$z_1, z_2 = 0, 1.$

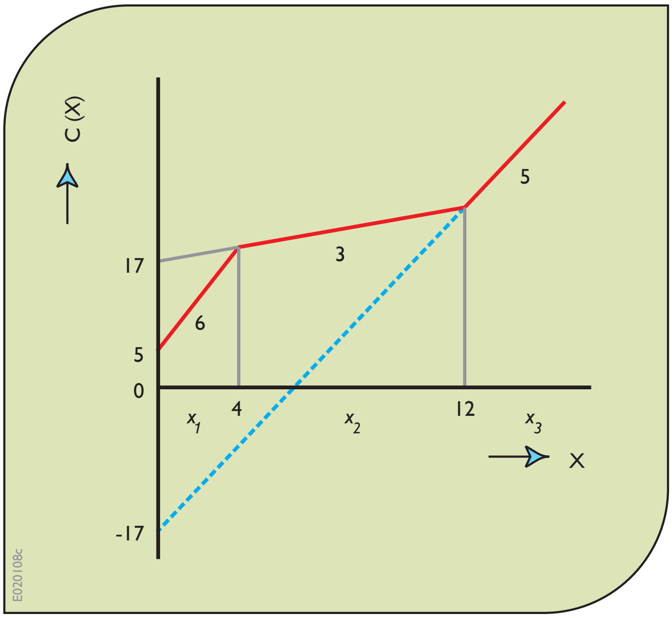


Maximize or Minimize $F(X)$

$$F(X) \cong (5z_1 + 6x_1) + (35z_2 + 3x_2) + (32z_3 - 2x_3) + 22z_4$$

Subject to:

$$x_1 + (4z_2 + x_2) + (12z_3 + x_3) + (17z_4 + x_4) = X;$$
$$x_1 \leq 4z_1; \quad x_2 \leq 8z_2; \quad x_3 \leq 5z_3; \quad x_4 \leq 99z_4;$$
$$\sum_s z_s = 1; \quad z_s = 0, 1 \ \forall s.$$




Maximize $C(X) \cong (5z_1 + 6x_1 + 3x_2) + (-17z_3 + 5x_3)$

Subject to: $(x_1 + x_2) + x_3 = X; z_1, z_3 = 0, 1.$

$$x_1 \leq 4z_1; \quad x_2 \leq 8z_1; \quad x_3 \leq 99z_3; \quad z_1 + z_3 = 1;$$

Minimize $C(X) \cong (5z_1 + 6x_1) + (17z_2 + 3x_2 + 5x_3)$

Subject to: $x_1 + (4z_2 + x_2 + x_3) = X; \quad z_1, z_2 = 0, 1.$

$$x_1 \leq 4z_1; \quad x_2 \leq 12z_2; \quad x_3 \leq 99z_2; \quad z_1 + z_2 \leq 1$$

## 6. A Brief Review

Before proceeding to some other optimization and simulation methods in the following chapters, it may be useful to review the topics covered so far. The focus has been on model development as well as model solution. Several types of water resources planning and management problems have been used to illustrate model development and solution processes. Like their real-world counterparts, the example problems all had multiple unknown decision-variables and multiple constraints. Also like their real-world counterparts, there are multiple feasible solutions to each of these problems. Hence, the task is to find the best solution, or a number of near-best solutions. Each solution must satisfy all the constraints.

Constraints can reflect physical conditions, environmental regulations and/or social or economic targets. Especially with respect to environmental or social conditions and goals, it is often a matter of judgement to decide what is considered an objective that is to be minimized or maximized and what is considered a constraint that has to be met.

Except for relatively simple problems, the use of these optimization models and methods is primarily for reducing the number of alternatives that need to be further analysed and evaluated using simulation methods. Optimization is generally used for preliminary screening – eliminating inferior alternatives before more detailed analyses are carried out. Presented were some approaches to preliminary screening involving calculus-based Lagrange multiplier and non-linear programming methods, discrete dynamic programming methods, and linear programming methods. Each method has its strengths and limitations.

The example problems used to illustrate these modelling and model solution methods have been relatively simple. However, simple applications such as these can form the foundation of models of more complex problems, as will be shown in later chapters.

## 7. Reference

BELLMAN, R.E. 1957. *Dynamic programming*. Princeton, N.J., Princeton University Press.

### Additional References (Further Reading)

BASSSON, M.S.; ALLEN, R.B.; PEGRAM, G.G.S. and VAN ROOYEN, J.A. 1994. *Probabilistic management of water resource and hydropower systems*. Highlands Ranch, Colo., Water Resources Publications.

BEROGGI, G.E.G. 1999. *Decision modelling in policy management: an introduction to analytic concepts*. Boston, Mass., Kluwer Academic.

BLANCHARD, B.S. and FABRYCKY, W.J. 1998. *Systems engineering and analysis*, 3rd edn. Upper Saddle River, N.J., Prentice Hall.

BURAS, N. 1966. *Dynamic programming and water resources development, advances in hydroscience*, Vol. 3. New York, Academic Press.

DORFMAN, R.; JACOBY, H.D. and THOMAS, H.A. Jr. 1972. *Models for managing regional water quality*. Cambridge, Mass., Harvard University Press.

ESOGBUE, A.O. (ed.). 1989 *Dynamic programming for optimal water resources systems analysis*. Englewood Cliffs, N.J., Prentice Hall.

HALL, W.A. and DRACUP, J.A. 1970. *Water resources systems engineering*. New York, McGraw-Hill.

HILLIER, F.S. and LIEBERMAN, G.J. 1990. *Introduction to operations research*, 5th edn. New York, McGraw-Hill.

HILLIER, F.S. and LIEBERMAN, G.J. 1990. *Introduction to stochastic models in operations research*. New York, McGraw-Hill.

HUFSCHMIDT, M.M. and FIERING, M.B. 1966. *Simulation techniques for design of water-resource systems.* Cambridge, Mass., Harvard University Press.

KARAMOUZ, M.; ZINSSER, W.K. and SZIDAROVSZKY, F. 2003. *Water resources systems analysis*. Boca Raton, Fla., Lewis.

LOUCKS, D.P.; STEDINGER, J.S. and HAITH, D.A. 1981. *Water resource systems planning and analysis*. Englewood Cliffs, N.J., Prentice Hall.

MAASS, A.; HUFSCHMIDT, M.M.; DORFMAN, R.; THOMAS, H.A. Jr.; MARGLIN, S.A. and FAIR, G.M. 1962. *Design of water-resource systems*. Cambridge, Mass., Harvard University Press.

MAJOR, D.C. and LENTON, R.L. 1979. *Applied water resources systems planning*. Englewood Cliffs, N.J., Prentice Hall.

MAYS, L.W. and TUNG, Y.-K. 1992. *Hydrosystems engineering and management*. New York, McGraw-Hill.

NEWNAN, D.G.; LAVELLE, J.P. and ESCHENBACH, T.G. 2002. *Essentials of engineering economic analysis*, 2nd edn. New York, Oxford University Press.

RARDIN, R.L. 1998. *Optimization in operations research*. Upper Saddle River, N.J., Prentice Hall.

REVELLE, C. 1999. *Optimizing reservoir resources*. New York, John Wiley.

SOMLYODY, L. 1997. Use of optimization models in river basin water quality planning: WATERMATEX'97. In: M.B. Beck and P. Lessard (eds.), *Systems analysis and computing in water quality management: towards a new agenda*. London, IWA.

TAYLOR, B.W. III. 1999. *Introduction to management science*, 6th edn. Upper Saddle River, N.J., Prentice Hall.

WILLIS, R. and YEH, W.W.-G. 1987. *Groundwater systems planning and management*. Englewood Cliffs, N.J., Prentice Hall.

WINSTON, W.L. 1994. *Operations research: applications and algorithms*, 3rd edn. Boston, Mass., Duxbury.

WINSTON, W.L. and ALBRIGHT, S.C. 2001. *Practical management science*, 2nd edn. Pacific Grove, Calif., Duxbury.

WRIGHT, J.R. and HOUCK, M.H. 1995. Water resources planning and management. In: W.F. Chen (ed.), *Civil engineering handbook*, Chapter 37. Boca Raton, CRC.

WURBS, R.A. 1996. *Modelling and analysis of reservoir system operations*. Upper Saddle River, N.J., Prentice Hall.

# 5. Fuzzy Optimization

1. Fuzziness: An Introduction 135
   1.1. Fuzzy Membership Functions 135
   1.2. Membership Function Operations 136

2. Optimization in Fuzzy Environments 136

3. Fuzzy Sets for Water Allocation 138

4. Fuzzy Sets for Reservoir Storage and Release Targets 139

5. Fuzzy Sets for Water Quality Management 140

6. Summary 144

7. Additional References (Further Reading) 144

# 5 Fuzzy Optimization

**The precise quantification of many system performance criteria and parameter and decision variables is not always possible, nor is it always necessary. When the values of variables cannot be precisely specified, they are said to be uncertain or fuzzy. If the values are uncertain, probability distributions may be used to quantify them. Alternatively, if they are best described by qualitative adjectives, such as dry or wet, hot or cold, clean or dirty, and high or low, fuzzy membership functions can be used to quantify them. Both probability distributions and fuzzy membership functions of these uncertain or qualitative variables can be included in quantitative optimization models. This chapter introduces fuzzy optimization modelling, again for the preliminary screening of alternative water resources plans and management policies.**

## 1. Fuzziness: An Introduction

Large, small, pure, polluted, satisfactory, unsatisfactory, sufficient, insufficient, excellent, good, fair, poor and so on are words often used to describe various attributes or performance measures of water resources systems. These descriptors do not have 'crisp', well-defined boundaries that separate them from others. A particular mix of economic and environmental impacts may be *more acceptable* to some and *less acceptable* to others. Plan A is *better* than Plan B. The water quality and temperature is *good* for swimming. These qualitative, or 'fuzzy', statements convey information despite the imprecision of the italicized adjectives.

This chapter illustrates how fuzzy descriptors can be incorporated into optimization models of water resources systems. Before this can be done some definitions are needed.

### 1.1. Fuzzy Membership Functions

Consider a set $A$ of real or integer numbers ranging from say 18 to 25. Thus $A = [18, 25]$. In classical (crisp) set theory, any number $x$ is either in or not in the set $A$. The statement '$x$ belongs to $A$' is either true or false depending on the value of $x$. The set $A$ is referred to as a crisp set.If one is not able to say for certain whether or not any number $x$ is in the set, then the set $A$ could be referred to as *fuzzy*. The degree of truth attached to that statement is defined by a membership function. This function ranges from 0 (completely false) to 1 (completely true).

Consider the statement, 'The water temperature should be suitable for swimming'. Just what temperatures are suitable will depend on the person asked. It would be difficult for anyone to define precisely those temperatures that are suitable if it is understood that temperatures outside that range are absolutely not suitable.

A membership function defining the interval or range of water temperatures suitable for swimming is shown in Figure 5.1. Such functions may be defined on the basis of the responses of many potential swimmers. There is a zone of imprecision or disagreement at both ends of the range.

The form or shape of a membership function depends on the individual subjective feelings of the 'members' or individuals who are asked their opinions. To define this particular membership function, each individual $i$ could be asked to define his or her comfortable water temperature interval ($T_{1i}$, $T_{2i}$). The membership value associated with any temperature value $T$ equals the number of

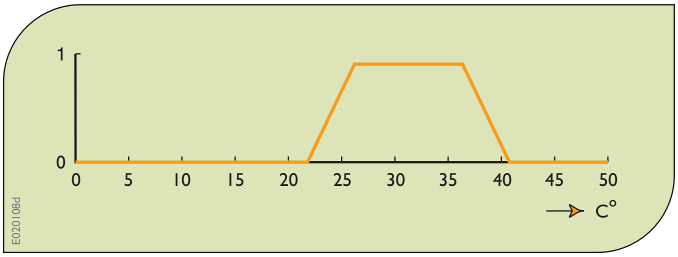



**Figure 5.1.** A fuzzy membership function for suitability of water temperature for swimming.

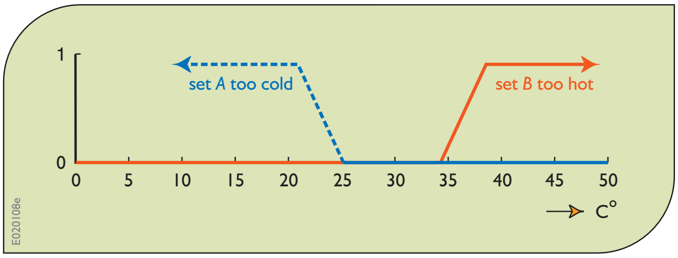



**Figure 5.2.** Two membership functions relating to swimming water temperature. Set *A* is the set defining the fraction of all individuals who think the water temperature is too cold, and Set *B* defines the fraction of all individuals who think the water temperature is too hot.

individuals who place that $T$ within their range ($T_{1i}$, $T_{2i}$), divided by the number of individual opinions obtained. The assignment of membership values is based on subjective judgements, but such judgements seem to be sufficient for much of human communication.

## 1.2. Membership Function Operations

Denote the membership function associated with a fuzzy set $A$ as $m_A(x)$. It defines the degree or extent to which any value of $x$ belongs to the set $A$. Now consider two fuzzy sets, $A$ and $B$. Set $A$ could be the range of temperatures that are considered too cold, and set $B$ could be the range of temperatures that are considered too hot. Assume these two sets are as shown in Figure 5.2.

The degree or extent that a value of $x$ belongs to either of two sets $A$ or $B$ is the maximum of the two individual membership function values. This union membership function is defined as:

$$m_{A\cup B}(x) = \text{maximum}(m_A(x), m_B(x)) \tag{5.1}$$

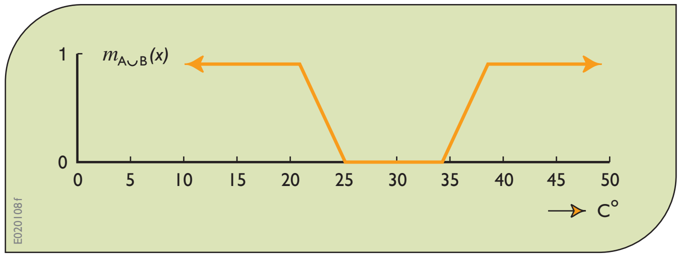



**Figure 5.3.** Membership function for water temperatures that are considered too cold or too hot.

This union set would represent the ranges of temperatures that are either too cold or too hot, as illustrated in Figure 5.3.

The degree or extent that a value of a variable $x$ is simultaneously in both sets $A$ and $B$ is the minimum of the two individual membership function values. This intersection membership function is defined as:

$$m_{A\cap B}(x) = \text{minimum}\ (m_A(x), m_B(x)) \tag{5.2}$$

This intersection set would define the range of temperatures that are considered both too cold and too hot. Of course it could be an empty set, as indeed it is in this case, based on the two membership functions shown in Figure 5.2. The minimum of either function for any value of $x$ is 0.

The complement of the membership function for fuzzy set $A$ is the membership function, $m_A^c(x)$, of $A^c$.

$$m_A^c(x) = 1 - m_A(x) \tag{5.3}$$

The complement of set $A$ (defined in Figure 5.2) would represent the range of temperatures considered not too cold for swimming. The complement of set $B$ (also defined in Figure 5.2) would represent the range of temperatures considered not too hot for swimming. The complement of the union set as shown in Figure 5.3 would be the range of temperatures considered just right. This complement set is the same as shown in Figure 5.1.

# 2. Optimization in Fuzzy Environments

Consider the problem of finding the maximum value of $x$ given that $x$ cannot exceed 11. This is written as:

$$\text{Maximize } U = x \tag{5.4}$$

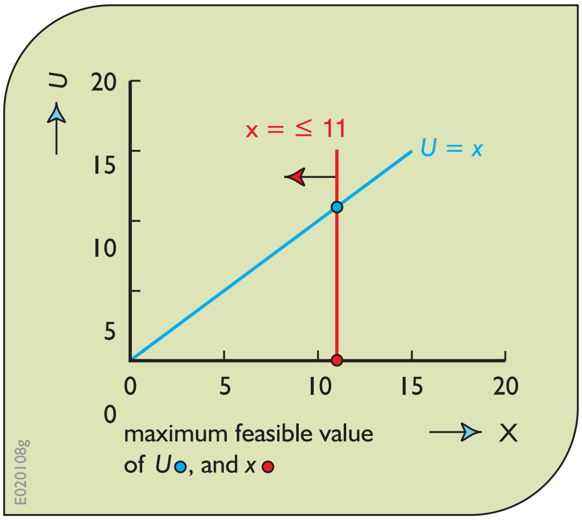



**Figure 5.4.** A plot of the crisp optimization problem defined by Equations 5.4 and 5.5.

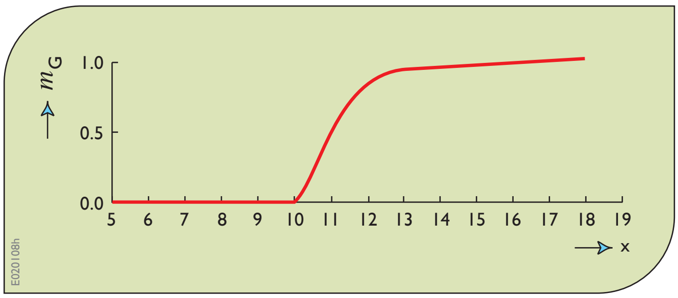



**Figure 5.5.** Membership function defining the fraction of individuals who think a particular value of *x* is 'substantially' greater than 10.

subject to:

$$x \leq 11 \tag{5.5}$$

The obvious optimal solution, $x = 11$, is shown in Figure 5.4.

Now suppose the objective is to obtain a value of $x$ substantially larger than 10 while making sure that the maximum value of $x$ should be in the vicinity of 11. This is no longer a crisp optimization problem; rather, it is a fuzzy one.

What is perceived to be substantially larger than 10 could be defined by a membership function, again representing the results of an opinion poll of what individuals think is substantially larger than 10. Suppose the membership function for this goal, $m_G(x)$, reflecting the results of such a poll, can be defined as:

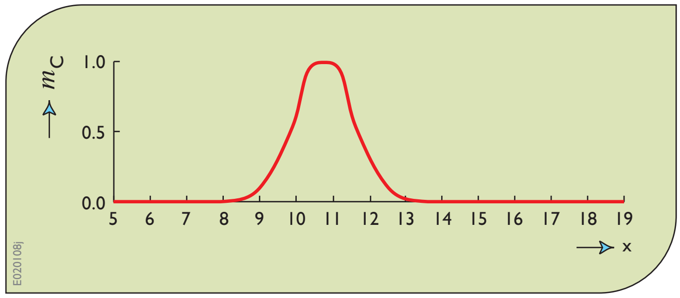



**Figure 5.6.** Membership function representing the vicinity of 11.

$$m_G(x) = 1/\{1 + [1/(x - 10)^2]\} \quad \text{if } x > 10$$
$$m_G(x) = 0 \text{ otherwise} \tag{5.6}$$

This function is shown in Figure 5.5.

The constraint on $x$ is that it 'should be in the vicinity of 11'. Suppose the results of a poll asking individuals to state what they consider to be in the vicinity of 11 results in the following constraint membership function, $m_C(x)$:

$$m_C(x) = 1/[1 + (x - 11)^4] \tag{5.7}$$

This membership function is shown in Figure 5.6.

Recall the objective is to obtain a value of $x$ substantially larger than 10 while making sure that the maximum value of $x$ should be in the vicinity of 11. In this fuzzy environment the objective is to maximize the extent to which $x$ exceeds 10 while keeping $x$ in the vicinity of 11. The solution can be viewed as finding the value of $x$ that maximizes the minimum values of both membership functions. Thus, we can define the intersection of both membership functions and find the value of $x$ that maximizes that intersection membership function.

The intersection membership function is:

$$\begin{aligned} m_D(x) &= \text{minimum}\{m_G(x), m_C(x)\} \\ &= \text{minimum}\{1/(1 + [1/(x - 10)^2]), \\ &\quad 1/(1 + (x - 11)^4)\} \quad \text{if } x > 10 \\ &= 0 \text{ otherwise} \end{aligned} \tag{5.8}$$

This intersection set, and the value of $x$ that maximizes its value, is shown in Figure 5.7.

This fuzzy decision is the value of $x$ that maximizes the intersection membership function $m_D(x)$, or equivalently:

$$\text{Maximize } m_D(x) = \max \min\{m_G(x), m_C(x)\} \tag{5.9}$$

Using LINGO®, the optimal solution is $x = 11.75$ and $m_D(x) = 0.755$.

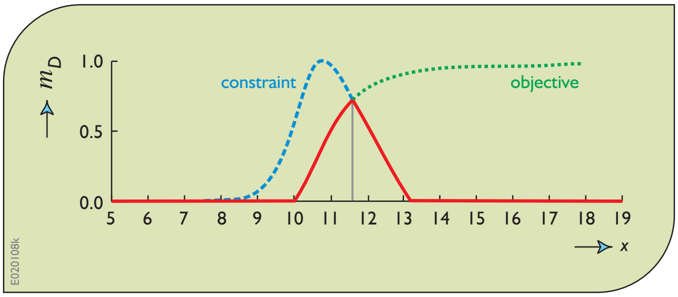



**Figure 5.7.** The intersection membership function and the value of *x* that represents a fuzzy optimal decision.

## 3. Fuzzy Sets for Water Allocation

Next consider the application of fuzzy modelling to the water allocation problem illustrated in Figure 5.8.

Assume, as in the previous uses of this example problem, the problem is to find the allocations of water to each firm that maximize the total benefits $TB(\mathbf{X})$:

$$\text{Maximize } TB(\mathbf{X}) = (6x_1 - x_1^2) + (7x_2 - 1.5x_2^2) + (8x_3 - 0.5x_3^2) \tag{5.10}$$

These allocations cannot exceed the amount of water available, $Q$, less any that must remain in the river, $R$. Assuming the available flow for allocations, $Q - R$, is 6, the crisp optimization problem is to maximize Equation 5.10 subject to the resource constraint:

$$x_1 + x_2 + x_3 \leq 6 \tag{5.11}$$

The optimal solution is $x_1 = 1$, $x_2 = 1$, and $x_3 = 4$ as previously obtained in Chapter 4 using several different optimization methods. The maximum total benefits, $TB(\mathbf{X})$, from Equation 5.10, equal 34.5.

To create a fuzzy equivalent of this crisp model, the objective can be expressed as a membership function of the set of all possible objective values. The higher the objective value the greater the membership function value. Since membership functions range from 0 to 1, the objective needs to be scaled so that it also ranges from 0 to 1.

The highest value of the objective occurs when there is sufficient water to maximize each firm's benefits. This unconstrained solution would result in a total benefit of 49.17 and this happens when $x_1 = 3$, $x_2 = 2.33$, and $x_3 = 8$. Thus, the objective membership function can be expressed by:

$$m(\mathbf{X}) = [(6x_1 - x_1^2) + (7x_2 - 1.5x_2^2) + (8x_3 - 0.5x_3^2)]/49.17 \tag{5.12}$$

It is obvious that the two functions (Equations 5.10 and 5.12) are equivalent. However, the goal of maximizing objective function 5.10 is changed to that of maximizing the degree of reaching the objective target. The optimization problem becomes:

$$\text{maximize } m(\mathbf{X}) = [(6x_1 - x_1^2) + (7x_2 - 1.5x_2^2) + (8x_3 - 0.5x_3^2)]/49.17$$

subject to:

$$x_1 + x_2 + x_3 \leq 6 \tag{5.13}$$

The optimal solution of (5.13) is the same as (5.10 and 5.11). The optimal degree of satisfaction is $m(\mathbf{X}) = 0.70$.

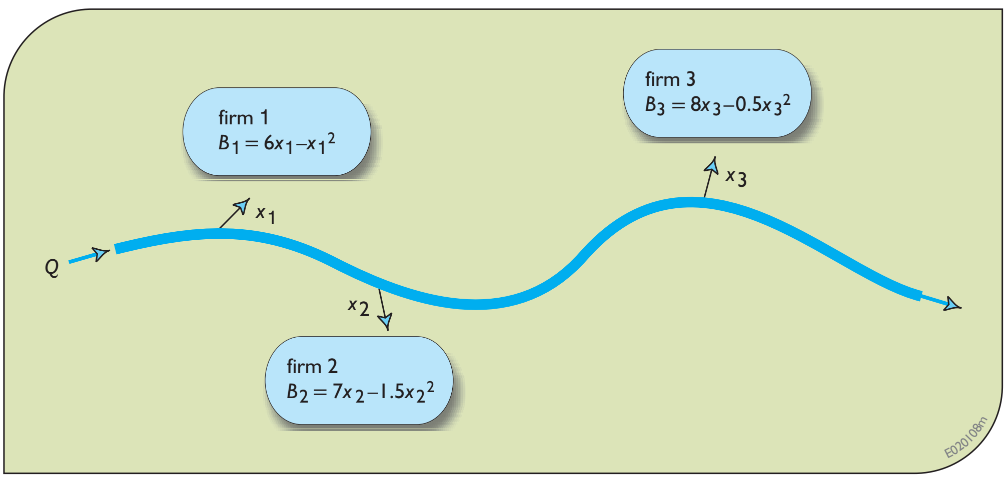



**Figure 5.8.** Three water-consuming firms *i* obtain benefits $B_i$ from their allocations $x_i$ of water from a river whose flow is *Q*.


**Figure 5.9. Membership function for 'about 6 units more or less'.**

Next, assume the amount of resources available to be allocated is limited to 'about 6 units more or less', which is a fuzzy constraint. Assume the membership function describing this constraint is defined by Equation 5.14 and is shown in Figure 5.9.

$$m_c(\mathbf{X}) = 1 \qquad \text{if } x_1 + x_2 + x_3 \leq 5$$
$$m_c(\mathbf{X}) = [7 - (x_1 + x_2 + x_3)]/2 \qquad \text{if } 5 \leq x_1 + x_2 + x_3 \leq 7$$
$$m_c(\mathbf{X}) = 0 \qquad \text{if } x_1 + x_2 + x_3 \geq 7 \qquad (5.14)$$

The fuzzy optimization problem becomes:
Maximize minimum ($m_G(\mathbf{X})$, $m_C(\mathbf{X})$)
subject to:

$$m_G(\mathbf{X}) = \left[\left(6x_1 - x_1^2\right) + \left(7x_2 - 1.5x_2^2\right) + \left(8x_3 - 0.5x_3^2\right)\right]/49.17$$
$$m_C(\mathbf{X}) = [7 - (x_1 + x_2 + x_3)]/2 \qquad (5.15)$$

Solving (5.15) using LINGO® to find the maximum of a lower bound on each of the two objectives, the optimal fuzzy decisions are $x_1 = 0.91$, $x_2 = 0.94$, $x_3 = 3.81$, $m(\mathbf{X}) = 0.67$, and the total net benefit, Equation 5.10, is $TB(\mathbf{X}) = 33.1$. Compare this with the crisp solution of $x_1 = 1$, $x_2 = 1$, $x_3 = 4$, and the total net benefit of 34.5.

## 4. Fuzzy Sets for Reservoir Storage and Release Targets

Consider the problem of trying to identify a reservoir storage volume target, $T^S$, for the planning of recreation facilities given a known minimum release target, $T^R$, and reservoir capacity $K$. Assume, in this simple example, these known release and unknown storage targets must apply in each of the three seasons in a year. The objective will be to find the highest value of the storage target, $T^S$, that minimizes the sum of squared deviations from actual storage volumes and releases less than the minimum release target.

| variable | value | remarks |
|---|---|---|
| $T^S$ | 15.6 | target storage for each period |
| $S_1$ | 19.4 | reservoir storage volume at beginning of period 1 |
| $S_2$ | 7.5 | reservoir storage volume at beginning of period 2 |
| $S_3$ | 20.0 | reservoir storage volume at beginning of period 3 |
| $R_1$ | 14.4 | reservoir release during period 1 |
| $R_2$ | 27.5 | reservoir release during period 2 |
| $R_3$ | 18.1 | reservoir release during period 3 |

E020827q

**Table 5.1. The LINGO® solution to the reservoir optimization problem.**

Given a sequence of inflows, $Q_t$, the optimization model is:

$$\text{Minimize } D = \sum_t^3 \left[(T^S - S_t)^2 + DR_t^2\right] - 0.001T^S \qquad (5.16)$$

subject to:

$$S_t + Q_t - R_t = S_{t+1} \quad t = 1, 2, 3; \quad \text{if } t = 3, t + 1 = 1 \qquad (5.17)$$
$$S_t \leq K \qquad t = 1, 2, 3 \qquad (5.18)$$
$$R_t \geq T^R - DR_t \quad t = 1, 2, 3 \qquad (5.19)$$

Assume $K = 20$, $T^R = 25$ and the inflows $Q_t$ are 5, 50 and 20 for periods $t = 1$, 2 and 3. The optimal solution, yielding an objective value of 184.4, obtained by LINGO® is listed in Table 5.1.

Now consider changing the objective function into maximizing the weighted degrees of 'satisfying' the reservoir storage volume and release targets.

$$\text{Maximize } \sum_t (w_S m_{St} + w_R m_{Rt}) \qquad (5.20)$$

where $w_S$ and $w_R$ are weights indicating the relative importance of storage volume targets and release targets respectively. The variables $m_{St}$ are the degrees of satisfying storage volume target in the three periods $t$, expressed by Equation 5.21. The variables $m_{Rt}$ are the degrees of satisfying release target in periods $t$, expressed by Equation 5.22.

$$m_S = \begin{cases} S_t/T^S & \text{for } S_t \leq T^S \\ (K-S_t)/(K-T^S) & \text{for } T^S \leq S_t \end{cases} \qquad (5.21)$$

$$m_R = \begin{cases} R_t/T^R & \text{for } R_t \leq T^R \\ 1 & \text{for } R_t > T^R \end{cases} \qquad (5.22)$$

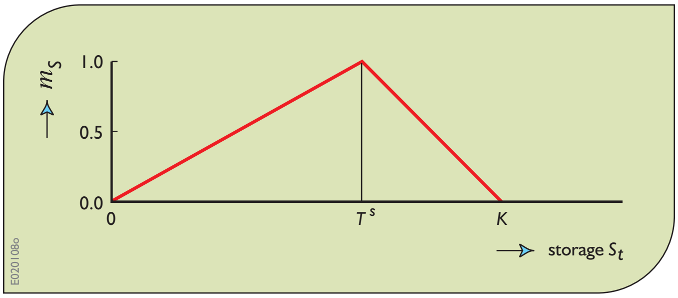



**Figure 5.10.** Membership function for storage volumes.

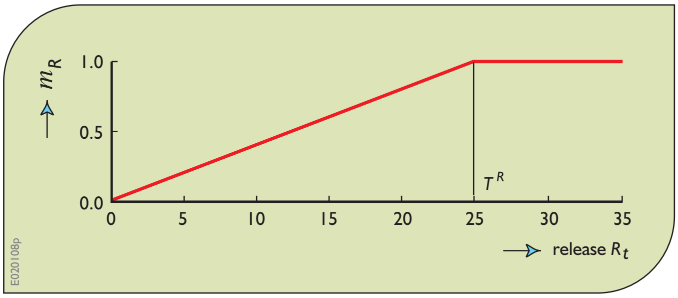



**Figure 5.11.** Membership function for releases.

### Box 5.1. Reservoir model written for solution by LINGO®

E020903d

```
SETS:
PERIODS /1..3/: I, R, m, ms, mr , s1, s2, ms1, ms2;
NUMBERS /1..4/: S;
ENDSETS
!*** OBJECTIVE ***; max = degree + 0.001*TS;
!Initial conditions; s(1) = s(TN + 1);
!Total degree of satisfaction; degree = @SUM(PERIODS(t): m(t));
!Weighted degree in period t; @FOR (PERIODS(t):
m(t) = ws*ms(t) + wr*mr(t);
S(t) = s1(t) + s2(t);
s1(t) < TS ;  s2(t) < K – TS  ;
!ms(t) = (s1(t)/TS) – (s2(t)/(K –TS)) =  rewritten in case dividing by 0;
ms1(t)*TS = s1(t);  ms2(t)*(K –TS) = s2(t);  ms(t) = ms1(t) – ms2(t);
 mr(t) < R(t)/TR ;   mr(t) < 1;  S(t+1) = S(t) + I(t) – R(t););
DATA:
TN = 3; K = 20; ws = ?; wr = ?;    I = 5, 50, 20;  TR = 25;
ENDDATA
```

Equations 5.21 and 5.22 are shown in Figures 5.10 and 5.11, respectively.

This optimization problem written for solution using LINGO® is as shown in Box 5.1.

Given weights $w_S = 0.4$ and $w_R = 0.6$, the optimal solution obtained from solving the model shown in Box 5.1 using LINGO® is listed in Table 5.2.

| variable | value | remarks |
|---|---|---|
| **degree** | 2.48 | total weighted sum membership function values |
| $T^s$ | 20.00 | target storage volume |
| $S_1$ | 20.00 | storage volume at beginning of period 1 |
| $S_2$ | 0.00 | storage volume at beginning of period 2 |
| $S_3$ | 20.00 | storage volume at beginning of period 3 |
| $R_1$ | 25.00 | reservoir release in period 1 |
| $R_2$ | 30.00 | reservoir release in period 2 |
| $R_3$ | 20.00 | reservoir release in period 3 |
| $M_1$ | 1.00 | sum weighted membership values period 1 |
| $M_2$ | 0.60 | sum weighted membership values period 2 |
| $M_3$ | 0.88 | sum weighted membership values period 3 |
| $M_1^s$ | 1.00 | storage volume membership value period 1 |
| $M_2^s$ | 0.00 | storage volume membership value period 2 |
| $M_3^s$ | 1.00 | storage volume membership value period 3 |
| $M_1^R$ | 1.00 | reservoir release membership value period 1 |
| $M_2^R$ | 1.00 | reservoir release membership value period 2 |
| $M_3^R$ | 0.80 | reservoir release membership value period 3 |

E020827s

**Table 5.2.** Solution of fuzzy model for reservoir storage volumes and releases based on objective 5.20.

If the objective Equation 5.20 is changed to one of maximizing the minimum membership function value, the objective becomes:

$$\text{Maximize } m_{\min} = \text{maximize minimum } \{m_{St}, m_{Rt}\} \qquad (5.23)$$

A common lower bound is set on each membership function, $m_{St}$ and $m_{Rt}$, and this variable is maximized. The optimal solution changes somewhat and is as shown in Table 5.3.

This solution differs from that shown in Table 5.2 primarily in the values of the membership functions. The target storage volume operating variable value, $T^S$, stays the same in this example.

## 5. Fuzzy Sets for Water Quality Management

Consider the stream pollution problem illustrated in Figure 5.12. The stream receives waste from sources located at Sites 1 and 2. Without some waste treatment at these sites, the pollutant concentrations at Sites 2 and 3 will exceed the maximum desired concentration. The problem is to find the level, $x_i$, of wastewater treatment (fraction of waste removed) at Sites $i = 1$ and 2 required to meet the quality standards at Sites 2 and 3 at a minimum total cost. The data used for the problem shown in Figure 5.12 are listed in Table 5.4.

| variable | value | remarks |
|---|---|---|
| MMF | 0.556 | minimum membership function value |
| $T^s$ | 20.00 | target storage volume |
| $S_1$ | 20.00 | storage volume at beginning of period 1 |
| $S_2$ | 11.11 | storage volume at beginning of period 2 |
| $S_3$ | 20.00 | storage volume at beginning of period 3 |
| $R_1$ | 13.88 | reservoir release in period 1 |
| $R_2$ | 41.11 | reservoir release in period 2 |
| $R_3$ | 20.00 | reservoir release in period 3 |
| $M_1^S$ | 0.556 | storage volume membership value period 1 |
| $M_2^S$ | 0.556 | storage volume membership value period 2 |
| $M_3^S$ | 0.800 | storage volume membership value period 3 |
| $M_1^R$ | 0.556 | reservoir release membership value period 1 |
| $M_2^R$ | 1.000 | reservoir release membership value period 2 |
| $M_3^R$ | 0.556 | reservoir release membership value period 3 |

**Table 5.3.** Optimal solution of reservoir operation model based on objective 5.23.

The crisp model for this problem, as discussed in the previous chapter, is:

$$\text{Minimize } C_1(x_1) + C_2(x_2) \tag{5.24}$$

subject to:

Water quality constraint at site 2:

$$[P_1Q_1 + W_1(1-x_1)]a_{12}/Q_2 \le P_2^{\max} \tag{5.25}$$

$$[(32)(10) + 250000(1-x_1)/86.4]\ 0.25/12 \le 20$$

which, when simplified, is: $x_1 \ge 0.78$

Water quality constraint at site 3:

$$\{[P_1Q_1 + W_1(1-x_1)]a_{13} + [W_2(1-x_2)]a_{23}\}/Q_3 \le P_3^{\max} \tag{5.26}$$

$$\{[(32)(10) + 250000(1-x_1)/86.4]\ 0.15 + [80000(1-x_2)/86.4]\ 0.60\}/13 \le 20$$

which, when simplified, is: $x_1 + 1.28x_2 \ge 1.79$

Restrictions on fractions of waste removal:

$$0 \le x_i \le 1.0 \quad \text{for sites } i = 1 \text{ and } 2 \tag{5.27}$$

For a wide range of reasonable costs, the optimal solution found using linear programming was 0.78 and 0.79, or essentially 80% removal efficiencies at Sites 1 and 2. Compare this solution with that of the following fuzzy model.

To develop a fuzzy version of this problem, suppose the maximum allowable pollutant concentrations in the stream at Sites 2 and 3 were expressed as 'about 20 mg/l or less'. Obtaining opinions of individuals of what

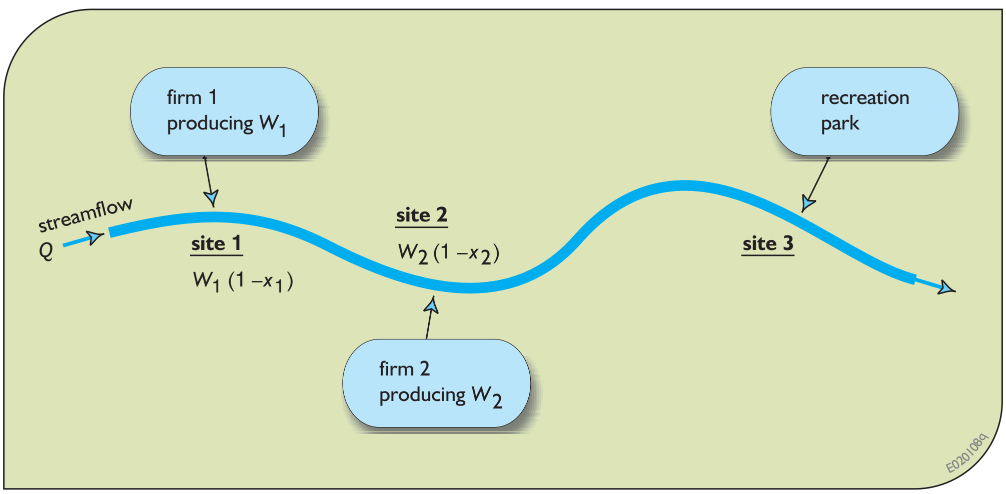



**Figure 5.12.** A stream pollution problem of finding the waste removal efficiencies $(x_1, x_2)$ that meet the stream quality standards at least cost.

**Table 5.4.** Parameter values selected for the water quality management problem illustrated in Figure 5.12.

| | parameter | unit | value | remark |
|---|---|---|---|---|
| flow | $Q_1$ | m$^3$/s | 10 | flow just upstream of site 1 |
| | $Q_2$ | m$^3$/s | 12 | flow just upstream of site 2 |
| | $Q_3$ | m$^3$/s | 13 | flow at park |
| waste | $W_1$ | kg/day | 250,000 | pollutant mass produced at site 1 |
| | $W_2$ | kg/day | 80,000 | pollutant mass produced at site 2 |
| pollutant conc. | $P_1$ | mg/l | 32 | concentration just upstream of site 1 |
| | $P_2$ | mg/l | 20 | maximum allowable concentration upstream of 2 |
| | $P_3$ | mg/l | 20 | maximum allowable concentration at site 3 |
| decay fraction | $a_{12}$ | -- | 0.25 | fraction of site 1 pollutant mass at site 2 |
| | $a_{13}$ | -- | 0.15 | fraction of site 1 pollutant mass at site 3 |
| | $a_{23}$ | -- | 0.60 | fraction of site 2 pollutant mass at site 2 |

E020827u

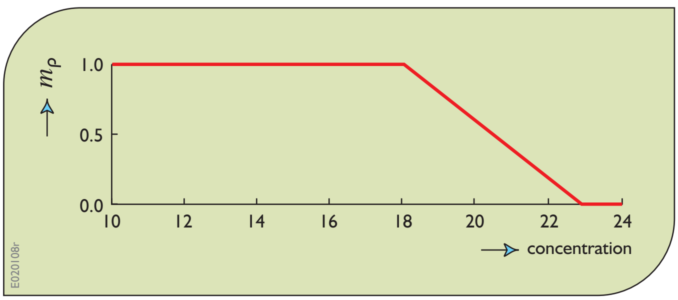



**Figure 5.13.** Membership function for 'about 20 mg/l or less'.

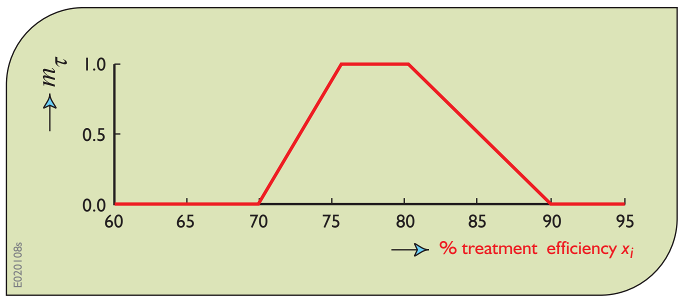



**Figure 5.14.** Membership function for best available treatment technology.

they consider to be '20 mg/l or less', a membership function can be defined. Assume it is as shown in Figure 5.13.

Next, assume that the government environmental agency expects each polluter to install best available technology (BAT) or to carry out best management practices (BMP) regardless of whether or not this is required to meet stream-quality standards. Asking experts just what BAT or BMP means with respect to treatment efficiencies could result in a variety of answers. These responses can be used to define membership functions for each of the two firms in this example. Assume these membership functions for both firms are as shown in Figure 5.14.

Finally, assume there is a third concern that has to do with equity. It is expected that no polluter should be required to treat at a much higher efficiency than any other polluter. A membership function defining just what differences are acceptable or equitable could quantify this concern. Assume such a membership function is as shown in Figure 5.15.

Considering each of these membership functions as objectives, a number of fuzzy optimization models can be defined. One is to find the treatment efficiencies that maximize the minimum value of each of these membership functions.

$$\text{Maximize } m = \max \min\{m_P, m_T, m_E\} \tag{5.28}$$

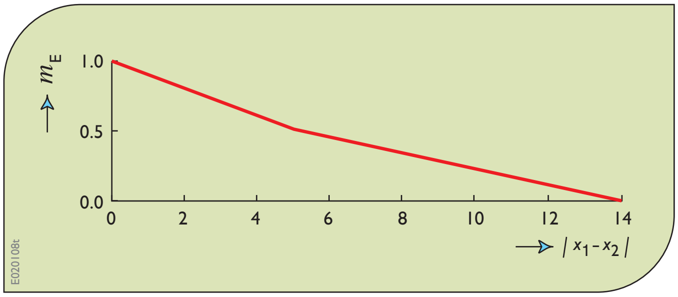



**Figure 5.15.** Equity membership function in terms of the absolute difference between the two treatment efficiencies.

If we assume that the pollutant concentrations at sites $j = 2$ and 3 will not exceed 23 mg/l, the pollutant concentration membership functions $m_{Pj}$ are:

$$m_{Pj} = 1 - p_{2j}/5 \tag{5.29}$$

The pollutant concentration at each site $j$ is the sum of two components:

$$p_j = p_{1j} + p_{2j} \tag{5.30}$$

where

$$p_{1j} \leq 18 \tag{5.31}$$

$$p_{2j} \leq (23 - 18) \tag{5.32}$$

If we assume the treatment plant efficiencies will be between 70 and 90% at both Sites $i = 1$ and 2, the treatment technology membership functions $m_{Ti}$ are:

$$m_{Ti} = (x_{2i}/0.05) - (x_{4i}/0.10) \tag{5.33}$$

and the treatment efficiencies are:

$$x_i = 0.70 + x_{2i} + x_{3i} + x_{4i} \tag{5.34}$$

where

$$x_{2i} \leq 0.05 \tag{5.35}$$

$$x_{3i} \leq 0.05 \tag{5.36}$$

$$x_{4i} \leq 0.10 \tag{5.37}$$

Finally, assuming the difference between treatment efficiencies will be no greater than 14, the equity membership function, $m_E$, is:

$$m_E = Z - (0.5/0.05)\, D_1 + 0.5(1 - Z) + (0.5/(0.14 - 0.05))\, D_2 \tag{5.38}$$

where

$$D_1 \leq 0.05Z \tag{5.39}$$

E020827v

| variable | value | remarks |
|---|---|---|
| $M$ | 0.93 | minimum membership value |
| $X_1$ | 0.81 | treatment efficiency at site 1 |
| $X_2$ | 0.81 | treatment efficiency at site 2 |
| $P_2$ | 18.28 | pollutant concentration just upstream of site 2 |
| $P_3$ | 18.36 | pollutant concentration just upstream of site 3 |
| $M^P{}_2$ | 0.94 | membership value for pollutant concentration site 2 |
| $M^P{}_3$ | 0.93 | membership value for pollutant concentration site 3 |
| $M^T{}_1$ | 0.93 | membership value for treatment level site 1 |
| $M^T{}_2$ | 0.93 | membership value for treatment level site 2 |
| $M^E$ | 1.00 | membership value for difference in treatment |

**Table 5.5.** Solution to fuzzy water quality management model Equations 5.28 to 5.46.

$$D_2 \leq (0.14 - 0.05)\,(1 - Z) \tag{5.40}$$

$$x_1 - x_2 = DP - DM \tag{5.41}$$

$$DP + DM = D_1 + 0.05(1 - Z) + D_2 \tag{5.42}$$

$Z$ is a binary 0, 1 variable. (5.43)

The remainder of the water quality model remains the same: Water quality constraint at site 2:

$$[P_1Q_1 + W_1(1 - x_1)]\, a_{12}/Q_2 = P_2 \tag{5.44}$$

$$[(32)(10) + 250000(1 - x_1)/86.4]\; 0.25/12 = P_2$$

Water quality constraint at site 3:

$$\{[P_1Q_1 + W_1(1 - x_1)]\, a_{13} + [W_2(1 - x_2)]\, a_{23}\}/Q_3 = P_3 \tag{5.45}$$

$$\{[(32)(10) + 250000(1 - x_1)/86.4]\; 0.15 + [80000(1 - x_2)/86.4]\; 0.60\}/13 = P_3$$

Restrictions on fractions of waste removal:

$$0 \leq x_i \leq 1.0 \text{ for sites } i = 1 \text{ and } 2. \tag{5.46}$$

Solving this fuzzy model using LINGO® yields the results shown in Table 5.5.

This solution confirms the assumptions made when constructing the representations of the membership functions in the model. It is also very similar to the least-cost solution found from solving the crisp LP model.

## 6. Summary

Optimization models incorporating fuzzy membership functions are sometimes appropriate when only qualitative statements are made when stating objectives and/or constraints of a particular water management problem or issue. This chapter has shown how fuzzy optimization can be applied to some simple example problems associated with water allocations, reservoir operation, and pollution control. This has been only an introduction. Those interested in more detailed explanations and applications may refer to any of the additional references listed in the next section.

## 7. Additional References (Further Reading)

BARDOSSY, A. and DUCKSTEIN, L. 1995. *Fuzzy rule-based modeling with applications to geophysical, biological, and engineering systems*. Boca Raton, Fla., CRC Press.

CHEN, S.Y. 1994. *Theory and application of fuzzy system decision-making*. Dalian, China, Dalian University of Technology Press. (In Chinese.).

KINDLER, J. 1992. Rationalizing water requirements with aid of fuzzy allocation model. *Journal of Water Resources Planning and Management*, ASCE, Vol. 118, No. 3, pp. 308–23.

KUNDZEWICZ, W. (ed.). 1995. *New uncertainty concepts in hydrology and water resources*. Cambridge, UK, Cambridge University Press.

LOOTSMA, F.A. 1997. *Fuzzy logic for planning and decision-making*. Boston, Mass., Kluwer Academic.

TERANO, T.; ASAI, K. and SUGENO, M. 1992. *Fuzzy systems theory and its application*. San Diego, Calif., Academic Press.

TILMANT, A.; VANCLOSSTER, M.; DUCKSTEIN, L. and PERSOONNS, E. 2002. Comparison of fuzzy and nonfuzzy optimal reservoir operating policies, *Journal of Water Resources Planning and Management*, ASCE, Vol. 128, No. 6, pp. 390–8.

ZHOU, H.-C. 2000. *Notes on fuzzy optimization and applications*. Dalian, China, Dalian University of Technology Press.

ZIMMERMANN, H.-J. 1987. *Fuzzy sets, decision-making, and expert systems*. Boston, Mass., Kluwer Academic.

# 6. Data-Based Models

1. Introduction 147

2. Artificial Neural Networks 148
   2.1. The Approach 148
   2.2. An Example 151
   2.3. Recurrent Neural Networks for the Modelling of Dynamic Hydrological Systems 153
   2.4. Some Applications 153
      2.4.1. RNN Emulation of a Sewerage System in the Netherlands 154
      2.4.2. Water Balance in Lake IJsselmeer 155

3. Genetic Algorithms 156
   3.1. The Approach 156
   3.2. Example Iterations 158

4. Genetic Programming 159

5. Data Mining 163
   5.1. Data Mining Methods 163

6. Conclusions 164

7. References 165

# 6 Data-Based Models

Data sets contain information, often much more than can be learned from just looking at plots of those data. Models based on observed input and output data help us abstract and gain new information and understanding from these data sets. They can also serve as substitutes for more process-based models in applications where computational speed is critical or where the underlying relationships are poorly understood. Models based on data range from the commonly used regression models to models based on, for example, evolutionary or biological processes. This chapter provides a brief introduction to these types of models. They are among a wide range of data-mining tools designed to identify and abstract information and/or relationships from large sets of data.

## 1. Introduction

Most optimization and simulation models used for water resources planning and management describe, in mathematical terms, the interactions and processes that take place among the various components of the system. These mechanistically or process-based models usually contain parameters whose values are determined from observed data during model calibration. These types of models are contrasted to what are typically called 'black-box' models, or statistical models. Statistical models do not describe physical processes. They attempt to convert inputs (e.g., rainfall, inflows to a reservoir, pollutants entering a wastewater treatment plant or discharged to a river) to outputs (e.g., runoff, reservoir releases, pollutant concentrations in the effluent of a treatment plant or downstream from a point of discharge to a river) using any mathematical equation or expression that does the job.

Regression equations, such as of the forms

$$\text{output} = a + b\,(\text{input}) \tag{6.1}$$

$$\text{output} = a + b\,(\text{input})^{c} \tag{6.2}$$

or

$$\text{output} = a + b_1\,(\text{input}_1)^{c1} + b_2\,(\text{input}_2)^{c2} \ldots \tag{6.3}$$

are examples of such data-based or statistical models. They depend on data consisting of both observed inputs and observed outputs for the estimation of the values of their parameters (a, b, c etc.) and for further refinement of their structure. They lack an explicit, well-defined representation of the processes involved in the transformation of inputs to outputs.

While these statistical models are better at interpolating within the range of data used to calibrate them, rather than extrapolating outside that range (as illustrated in Figure 6.1), many have proven quite successful in representing complex physical systems within those ranges.

Regression equations are examples of data-based modelling methods. Other examples of such methods are based on evolutionary principles and concepts. These are a class of probabilistic search procedures known as evolutionary algorithms (EAs). Such algorithms include genetic algorithms (GAs), genetic or evolutionary programming (GP or EP) and evolutionary strategy (ES). Each of these methods has many varieties but all use computational methods based on natural evolutionary processes and learning. Perhaps the most robust and hence the most common of these methods are genetic algorithms and all their varieties, and genetic programming. Alternatively, an extension of regression is artificial neural networks (ANN). The development and application of black-box statistical models like GP and ANNs emulate larger, deterministic, process-oriented models. Their use may be advantageous when for some reason a

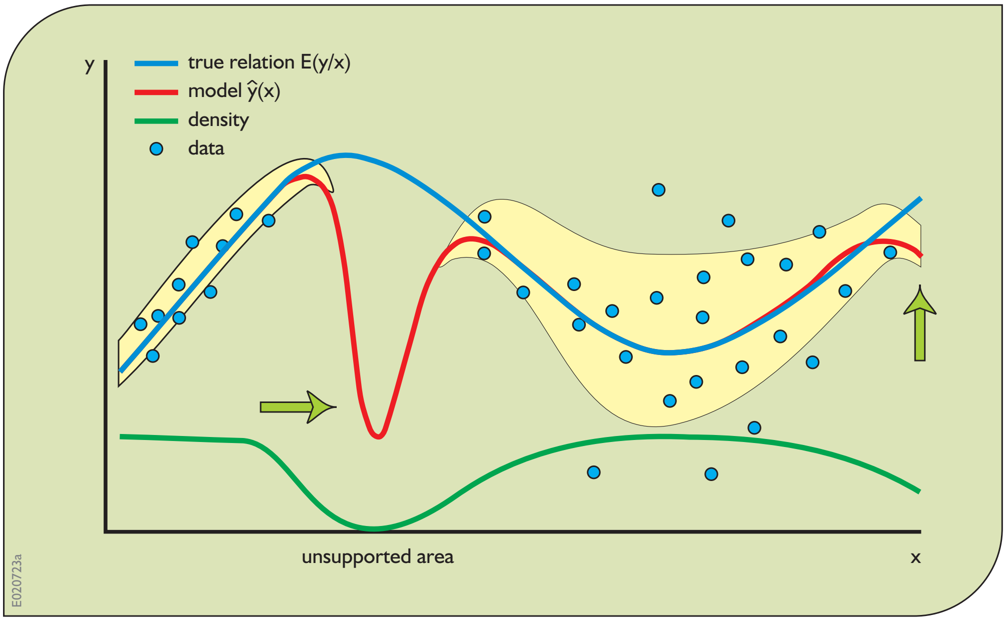



**Figure 6.1.** Data-driven models are able to estimate relatively accurately within their calibrated ranges, but not outside those ranges. The bottom curve represents the relative density of data used in model calibration. The arrows point to where the model does not predict well.

large number of model solutions must be obtained in a short period of time.

Examples of such situations where multiple solutions of a model must be obtained include sensitivity or uncertainty analysis, scenario evaluations, risk assessment, optimization, inverse modelling to obtain parameter values, and/or when model runs must be extremely fast, as for rapid assessment and decision support systems, real-time predictions/management/control, and so on. Examples of the use of data-driven models for model emulation are given in the following sections.

*Genetic algorithms* and *genetic programming* are automated, domain-independent methods for evolving solutions to existing models or for producing new models that emulate actual systems, such as rainfall–runoff relationships in a watershed, wastewater removal processes in a treatment plant or discharges of water from a system of natural lakes, each subject to random inputs.

Search methods such as genetic algorithms and genetic programming are inspired by our understanding of biology and natural evolution. They start initially with a number of randomly created values of the unknown variables or a number of black-box models, respectively. The variable values or structure of each of these models is progressively improved over a series of generations. The evolutionary search uses the Darwinian principal of 'survival of the fittest' and is patterned after biological operations including crossover (sexual recombination), mutation, gene duplication and gene deletion.

*Artificial neural networks* are distributed, adaptive, generally nonlinear networks built from many different processing elements (PEs) (Principe et al., 2000). Each processing element receives inputs from other processing elements and/or from itself. The inputs are scaled by adjustable parameters called weights. The processing elements sum all of these weighted inputs to produce an output that is a nonlinear (static) function of the sum. Learning (calibration) is accomplished by adjusting the weights. The weights are adjusted directly from the training data without any assumptions about the data's statistical distribution or other characteristics (Hagan et al., 1996; Hertz et al., 1991).

The following sections are intended to provide some background helpful to those who may be selecting one among all the available computer codes for implementing a genetic algorithm, genetic program or artificial neural network. These sections are largely based on Babovic and Bojkov (2001) and from material provided by Mynett and Van den Boogaard of Delft Hydraulics.

## 2. Artificial Neural Networks

### 2.1. The Approach

Before the development of digital computers, any information processing necessary for thinking and reasoning was carried out in our brains. Much of it still is. Brain-based information processing continues today and will

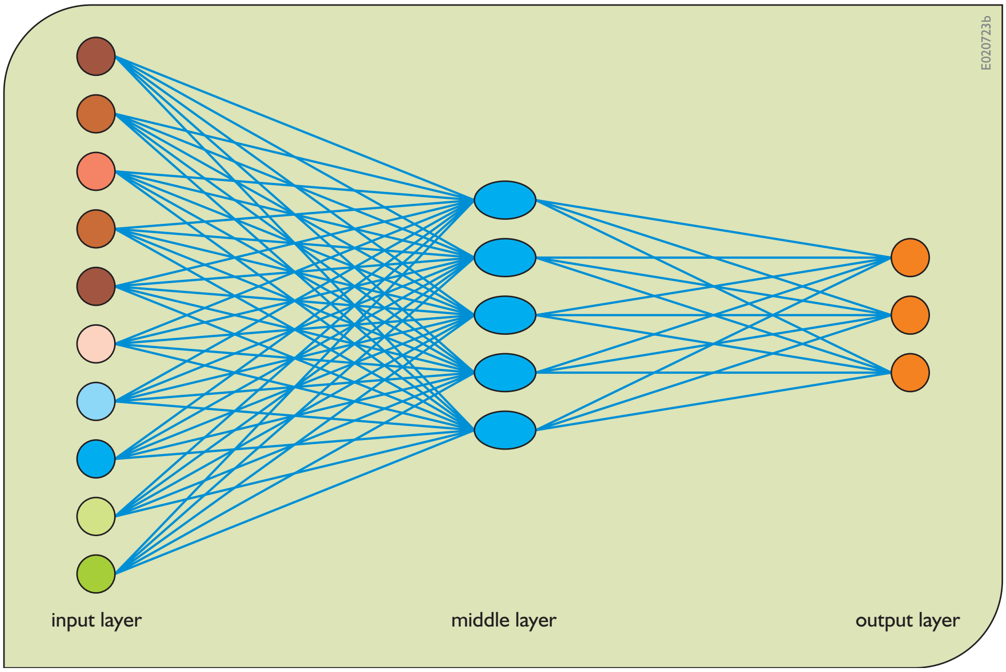



**Figure 6.2.** A typical multi-layer artificial neural network showing the input layer for ten different inputs, the hidden layer(s), and the output layer having three outputs.

continue in the future even given our continually improving digital processing technology capabilities. While recent developments in information technology (IT) have mastered and outperformed much of the information processing one can do just using brain power, IT has not mastered the reasoning power of our brains. Perhaps because of this, some computer scientists have been working on creating information processing devices that mimic the human brain. This has been termed *neurocomputing*. It operates with networks (ANNs) representing simplified models of the brain. In reality, it is just a more complex type of regression or statistical (black-box) model.

The basic structure of an ANN is shown in Figure 6.2. There are a number of input layer nodes on the left side of the figure and a number of output layer nodes on the right. The middle columns of nodes between these input and output nodes are called *hidden layers*. The number of hidden layers and the number of nodes in each layer are two of the design parameters of any ANN.

Most applications require networks that contain at least three types of layers:

- *The input layer* consists of nodes that receive an input from the external environment. These nodes do not perform any transformations upon the inputs but just send their weighted values to the nodes in the immediately adjacent, usually 'hidden,' layer.
- *The hidden layer(s)* consists of nodes that typically receive the transferred weighted inputs from the input layer or previous hidden layer, perform their transformations on it, and pass the output to the next adjacent layer, which can be another hidden layer or the output layer.
- *The output layer* consists of nodes that receive the hidden-layer output and send it to the user.

The ANN shown in Figure 6.2 has links only between nodes in immediately adjacent layers or columns and is often referred to as a multi-layer perceptron (MLP) network, or a feed-forward (FF) network. Other architectures of ANNs, which include recurrent neural networks (RNN), self-organizing feature maps (SOFMs), Hopfield networks, radial basis function (RBF) networks, support vector machines (SVMs) and the like, are described in more detail in other publications (for example, Haykin, 1999; Hertz et al., 1991).

Essentially, the strength (or weight) of the connection between adjacent nodes is a design parameter of the ANN. The output values $O_j$ that leave a node $j$ on each of its outgoing links are multiplied by a weight, $w_j$. The input $I_k$ to each node $k$ in each middle and output layer is the sum of each of its weighted inputs, $w_j\, O_j$ from all nodes $j$ providing inputs (linked) to node $k$.

$$\text{Input value to node } k\text{: } I_k = \sum_j w_j O_j \qquad (6.4)$$

Again, the sum in Equation 6.4 is over all nodes $j$ providing inputs to node $k$.

At each node $k$ of hidden and output layers, the input $I_k$ is an argument to a linear or non-linear function $f_k(I_k + \theta_k)$, which converts the input $I_k$ to output $O_k$. The variable $\theta_k$ represents a bias or threshold term that influences the horizontal offset of the function. This transformation can take on a variety of forms. A commonly used transformation is a sigmoid or logistic function as defined in Equation 6.5 and graphed in Figure 6.3.

$$O_k = 1/[1 + \exp\{-(I_k + \theta_k)\}] \tag{6.5}$$

The process of converting inputs to outputs at each hidden layer node is illustrated in Figure 6.4. The same process also happens at each output layer node.

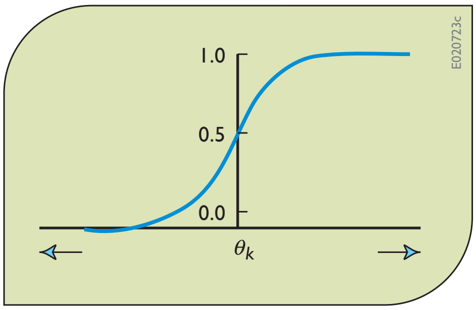



**Figure 6.3.** The sigmoid or logistic threshold function with threshold $\theta_k$.

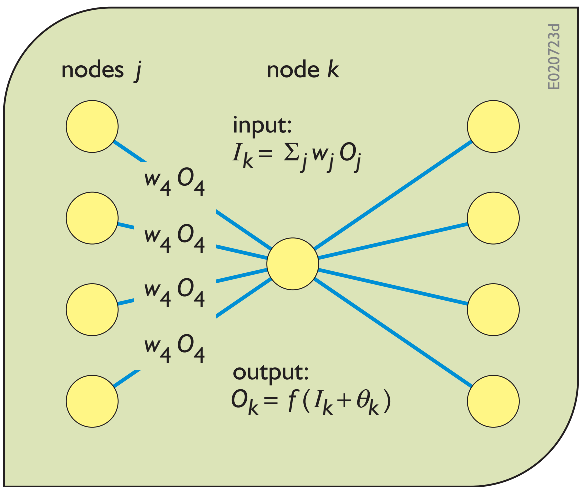



**Figure 6.4.** A hidden-layer node $k$ converting input values to an output value using a non-linear function $f$ (such as defined by Equation 6.5) in a multi-layer ANN.

The design issues in artificial neural networks are complex and are major concerns of ANN developers. The number of nodes in the input as well as in the output layer is usually predetermined from the problem to be solved. The number of nodes in each hidden layer and the number of hidden layers are calibration parameters that can be varied in experiments focused on getting the best fit of observed and predicted output data-based on the same input data. These design decisions, and most importantly the determination of the values of the weights and thresholds of each connection, are 'learned' during the 'training' of the ANN using predefined (or measured) sets of input and output data.

Some of the present-day ANN packages provide options for building networks. Most provide fixed network layers and nodes. The design of an ANN can have a significant impact on its data-processing capability.

There are two major connection topologies that define how data flows between the input, hidden and output nodes. These main categories are:

- *Feed-forward networks* in which the data flow through the network in one direction from the input layer to the output layer through the hidden layer(s). Each output value is based solely on the current set of inputs. In most networks, the nodes of one layer are fully connected to the nodes in the next layer (as shown in Figure 6.2); however, this is not a requirement of feed-forward networks.
- *Recurrent or feedback networks* in which, as their name suggests, the data flow not only in one direction but in the opposite direction as well for either a limited or a complete part of the network. In recurrent networks, information about past inputs is fed back into and mixed with inputs through recurrent (feedback) connections. The recurrent types of artificial neural networks are used when the answer is based on current data as well as on prior inputs.

Determining the best values of all the weights is called training the ANN. In a so-called supervised learning mode, the actual output of a neural network is compared to the desired output. Weights, which are usually randomly set to begin with, are then adjusted so that the next iteration will produce a closer match between the desired and the actual output. Various learning methods for weight adjustments try to minimize the differences or errors between observed and computed output data. Training consists of presenting input and output data to the network. These data are often referred to as training data. For each input provided to the network, the corresponding desired output set is provided as well.

The training phase can consume a lot of time. It is considered complete when the artificial neural network reaches a user-defined performance level. At this level the network has achieved the desired statistical accuracy as it produces the required outputs for a given sequence of inputs. When no further learning is judged necessary, the resulting weights are typically fixed for the application.

Once a supervised network performs well on the training data, it is important to see what it can do with data it has not seen before. If a system does not give a reasonable output for this test set, this means that the training period should continue. Indeed, this testing is critical to ensure that the network has learned the general patterns involved within an application and has not simply memorized a given set of data.

Smith (1993) suggests the following procedure for preparing and training an ANN:

1. Design a network.
2. Divide the data set into training, validation and testing subsets.
3. Train the network on the training data set.
4. Periodically stop the training and measure the error on the validation data set.
5. Save the weights of the network.
6. Repeat Steps 2, 3 and 4 until the error on the validation data set starts increasing. This is the moment where the overfitting has started.
7. Go back to the weights that produced the lowest error on the validation data set, and use these weights for the trained ANN.
8. Test the trained ANN using the testing data set. If it shows good performance use it. If not, redesign the network and repeat entire procedure from Step 3.

There is a wide selection of available neural network models. The most popular is probably the multi-layer feed-forward network, which is typically trained with static back propagation. They are easy to use, but they train slowly, and require considerable training data. In fact, the best generalization performance is produced if there are at least thirty times more training samples than network weights (Haykin, 1999). Adding local recurrent connections can reduce the required network size, making it less sensitive to noise, but it may get stuck on a solution that is inferior to what can be achieved.

## 2.2. An Example

To illustrate how an ANN might be developed, consider the simple problem of predicting a downstream pollutant concentration based on an upstream concentration and the streamflow. Twelve measurements of the streamflow quantity, velocity and pollutant concentrations at two sites (an upstream and a downstream site) are available. The travel times between the two measurement sites have been computed and these, plus the pollutant concentrations, are shown in Table 6.1.

Assume at first that the ANN structure consists of two input nodes, one for the upstream concentration and the

| travel time (days) | concentration upstream | concentration downstream |
|---|---|---|
| 2.0 | 20.0 | 6.0 |
| 2.0 | 15.0 | 4.5 |
| 1.5 | 30.0 | 12.2 |
| 1.0 | 20.0 | 11.0 |
| 0.5 | 20.0 | 14.8 |
| 1.0 | 15.0 | 8.2 |
| 0.5 | 30.0 | 22.2 |
| 1.5 | 25.0 | 10.2 |
| 1.5 | 15.0 | 6.1 |
| 2.0 | 30.0 | 9.0 |
| 1.0 | 30.0 | 16.5 |
| 0.5 | 25.0 | 18.5 |

E020723e

**Table 6.1.** Streamflow velocities and pollutant concentrations.

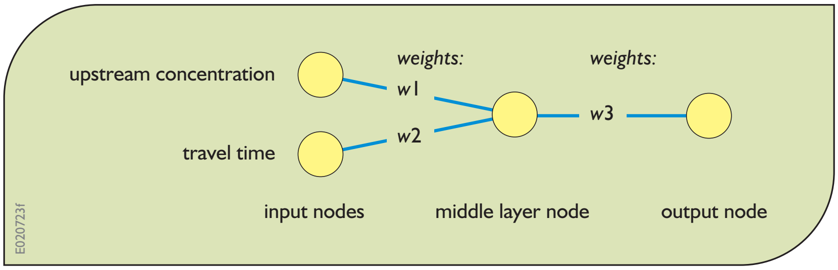



**Figure 6.5.** Initial ANN for example problem.

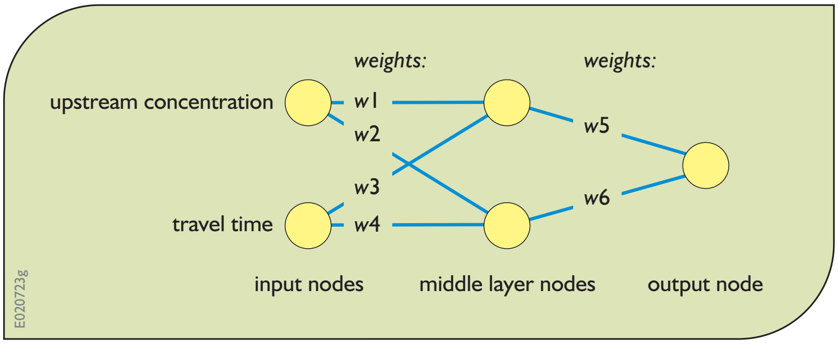



**Figure 6.6.** Modified ANN for example problem.

**Table 6.2.** Weights for each link of the ANN shown in Figure 6.6 based on six data sets from Table 6.1. All bias variables ($\theta_k$ in Equation 6.5) were 0.

| weights | value | weights | value |
|---|---|---|---|
| $w_1$ | 0.0 | $w_5$ | 8.1 |
| $w_2$ | 0.0 | $w_6$ | -2.8 |
| $w_3$ | -0.6 | | |
| $w_4$ | 3.9 | | |

E020723h

other for the travel time, a single hidden layer of only one node, and a single output node, the downstream concentration expressed as a fraction of the upstream concentration. This is shown in Figure 6.5.

The model output is the fraction of the upstream concentration that reaches the downstream site. That fraction can be any value from 0 to 1. Hence the sigmoid function (Equation 6.5) is applied at the middle node and at the output node. Using two or more data sets to train or calibrate this ANN (Figure 6.5) results in a poor fit as measured by the minimum sum of absolute deviations between calculated and measured concentration data. The more data samples used, the worse the fit. This structure is simply too simple. Hence, another node was added to the middle layer. This ANN is shown in Figure 6.6.

Using only half the data (six data sets) for training or calibration, the weights obtained provided a near perfect fit. The weights obtained are shown in Table 6.2.

Next the remaining six data sets were applied to the network with weights set to those values shown in Table 6.2. Again the sum of absolute deviations was essentially 0. Similar results were obtained with increasing numbers of data sets.

The weights in Table 6.2 point out something water quality modellers typically assume, and that is that the fraction of the upstream pollutant concentration that reaches the downstream site is independent of the actual upstream concentration (see Chapter 12). This ANN could have had only one input node, namely that for travel time. This conforms to the typical first order decay function:

$$\text{Fraction of pollutant concentration downstream per unit concentration upstream} = \exp\{-k^*(\text{travel time})\} \quad (6.6)$$

where the parameter $k$ is the decay rate constant (travel time units$^{-1}$).

## 2.3. Recurrent Neural Networks for the Modelling of Dynamic Hydrological Systems

The flexibility, efficiency, speed and emulation capacity of data-based models are particularly relevant within the fields of hydrology, meteorology and hydraulics, where one often has to model dynamic systems and processes that evolve in time. The previous illustration of an ANN was typical of many static ANNs. In training and/or applications the input–output patterns could be processed 'independently' in the sense that an input pattern is not affected by the ANN's response to previous input patterns. The serial order of the input–output patterns does not really matter.

This situation is the same as with the more commonly applied linear or non-linear regression and interpolation techniques where curves or surfaces are fitted to a given set of data points. In fact, standard ANNs are universal function approximators (Cybenko, 1989; Hornik et al., 1989) and, as such, are static non-linear regression models without explicit consideration of time, memory, and time evolution and/or interaction of input–output patterns.

For dynamic systems, the dependencies and/or interaction of input–output patterns cannot be ignored. This is perhaps most recognizable from the state–space property of dynamic models. In a dynamic state–space model, the system's state in time $t$ is based upon the system's state at time $t-1$ (and/or more previous time steps) and external forcings on the system. (See, for example, the applications of dynamic programming to reservoir operation in Chapter 4.)

This state–space property of dynamic models can be included within an ANN to create data-based models better suited for non-linear modelling and the analysis of dynamic systems and time series. This extension leads to the use of recurrent neural networks. These models are equipped with a time propagation mechanism that involves feedback of computed outputs rather than feedback of observed outputs, as shown in Figure 6.7. More detail concerning RNNs can be found in Van den Boogaard et al. (2000).

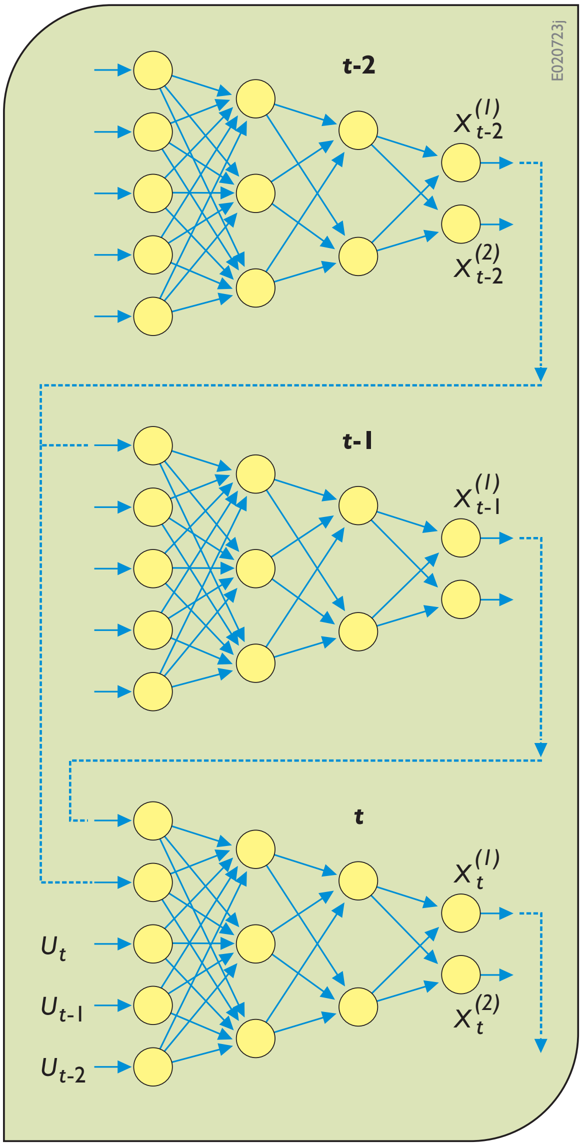



**Figure 6.7.** Illustration of the architecture of a recurrent neural network (RNN) where the inputs include external data $U$ as well as computed data $X$ from previous periods $t$.

## 2.4. Some Applications

Here two hydrological RNN applications are presented. The first case applies a RNN for the emulation of a dynamic conceptual model of a sewerage system. The second case applies a RNN to the modelling of the water balance in Lake IJsselmeer in the Netherlands.

### 2.4.1. RNN Emulation of a Sewerage System in the Netherlands

A major problem in sewerage systems is overflow during severe rainstorm events. An overflow occurs when the capacity of the sewerage system or treatment plants is exceeded during a rainfall event. Overflow devices are often no more than a storage chamber with a weir as an element for flow control. An overflow results when the upstream water depth exceeds the weir's crest.

Approximately 90% of all sewerage systems in the Netherlands are combined systems, spilling diluted sewage into open water systems during extreme storm events. For most towns, these systems have mild slopes, if any, which implies that the system forms an interconnected network with many loops. When all street sewer lines are included, the resultant networks may have thousands or even tens of thousands of pipes.

For the simulation of such complex sewerage systems, models based on the numerical solution of the Saint Venant equations can be used. One such deterministic model is the SOBEK-URBAN model jointly developed by three organizations in the Netherlands: WL | Delft Hydraulics, RIZA and DHV. As one might expect, this is the principal tool used in the Netherlands for the simulation of sewerage systems (Stelling, 2000). This modelling platform provides an integrated approach for a 1-D simulation of processes in rivers, sewers and drainage systems. Flows in the pipe network and the receiving waters can easily be combined. Another well-known deterministic model applicable for sewerage systems is the INFOWORKS model of Wallingford Software in the UK.

Over the past years, legislation regarding sewage spilling to open water systems has placed more and more constraints on the amounts and frequency of overflows allowed during storm events. To assess whether or not a particular sewerage system meets the current legislation, simulations over a ten-year period are required. In the near future, legislation may require that this period be extended to about twenty-five years.

Simulations of sewerage systems are performed on the basis of recorded historical series of rainstorm events. For a given layout, or for proposed rehabilitation designs (with, say, additional storage capacities), this historical set of storm events must be simulated. For a time period of ten or twenty-five years this will typically involve at least several hundred events where sewer overflows result from storms.

For large systems containing several thousands of pipes, such simulations are, and will continue to be, computationally burdensome despite increased computer speed and memory capacity. Even when restricted to subsets of events with potential overflows, the computational effort will still be large, especially when alternative designs must be compared. For this reason, RNNs as fast model emulators and/or model reduction devices have been developed (Price et al., 1998; Proaño et al., 1998).

The external input consists of a rainfall time series during a storm event. The output of the RNN is a time series of overflow discharges at one or more overflow structures. This output must be zero when no overflow occurs. The RNN is calibrated on the basis of an ensemble of rainfall–overflow combinations with the overflow time series generated by a numerical model. The calibration (and verification) set includes rainstorm events with overflow as well as events without overflows. In this way, the output time series of a few nodes of the numerical model are emulated, rather than the complete state of the sewerage system.

An important question is whether or not the emulation can be based on a limited subset of the large 'original' ensemble of events. Otherwise the input–output of virtually all events must be pre-computed with the numerical model, and the desired reduction of computational cost will not be achieved. Apart from that, emulation would no longer make sense as the frequency and quantity of overflows could then be established directly from the numerical model's predictions.

This model emulation/reduction was applied to the sewerage system of Maartensdijk, a town with about 10,000 inhabitants in the central Netherlands. Based on the system emptying time and the storage/pump capacities, 200 potential overflow events were selected from a rainfall series of ten years. For all forty-four events of the calibration and verification set, a simulation with a SOBEK-URBAN model of the Maartensdijk sewerage system was carried out, providing time series of water depths and discharges at three overflow structures in the system. With these depth and discharge series, a RNN-model was calibrated and verified. The weights of the RNN were optimized for all training events simultaneously and not for each event separately.

This study showed that the water depth rather than the discharge determines the state of the system. In fact, the discharges are a function of the water depths (rating curve), but this relation is not one to one (for example: the discharge is zero for any water depth below the weir's crest). Therefore, discharges cannot be used for the representation of the system state. This aspect nicely illustrates that within black-box procedures too, physical principles or system knowledge must often be used to guarantee a meaningful model set-up.

After several training experiments, a proper architecture of the RNN was identified and accurate emulations of the SOBEK-URBAN model were obtained. Validation tests on events not contained in the training and verification set demonstrated the RNN's capability to emulate the deterministic model.


**Figure 6.8.** Plan view of Lake Ijsselmeer.

## 2.4.2. Water Balance in Lake IJsselmeer

The most important factors that affect the water balance of Lake IJsselmeer are the inflow of the River IJssel and outflows at the sluices at Den Oever and Kornwerderzand. Figure 6.8 is a map of the area with the positions of the River IJssel and the sluices denoted. The sluices form an interface of the lake to the tidal Dutch Wadden Sea and are used to discharge excess water. This spilling is only possible during low tides when the outside water levels are lower than the water levels within the lake. The spilled volumes are determined by the difference of these water levels. The inner water level can be significantly affected by wind.

An RNN was prepared for the dynamic modelling of the water levels of Lake Ijsselmeer (Van den Boogaard et al., 1998). The model included five external system forcings. These were the discharge of the River IJssel, the north–south and east–west components of the wind, and the outside tidal water levels at the sluices of Kornwerderzand and Den Oever. The (recurrent) output consists of the water level of Lake IJsselmeer. The time step in the model is one day, adopted from the sampling rate of the observed discharges of the River IJssel and the lake's water levels. For synchronization, the time series of the wind and the outer tidal water levels are also represented by daily samples.

The discharges of the River IJssel are a major source of uncertainty. They were not directly observed but on the basis of empirical relations using water-level registrations of the River Rhine at Lobith near the Dutch–German border. These relations may involve errors, especially for winter seasons when water levels and/or discharges are large or even extreme. In all the other external forcings, such as the wind, large uncertainties may also be present.

For applications under operational or real-time conditions (e.g., short to medium-term water-level forecasting) these uncertainties can be taken into account more explicitly by modelling the uncertainties in a statistical sense and applying online data assimilation techniques to improve the model's ability to forecast. Van den Boogaard et al., (2000) present applications of online data assimilation to this situation. The results from this RNN model are probably as good as, if not better than, could be expected from a much more detailed hydraulic process-based model.

For other applications of artificial neural networks applied to problems in meteorology, oceanography, hydraulics, hydrology, and ecology, see Wüst (1995), Minns (1996), Minns and Hall (1996), Scardi (1996), Abrahart et al., (2004), Recknagel et al., (1997), Clair and Ehrman (1998), Hsieh and Tang (1998), Lange (1998), Sanchez et al., (1998), Shen et al., (1998), Van Gent and

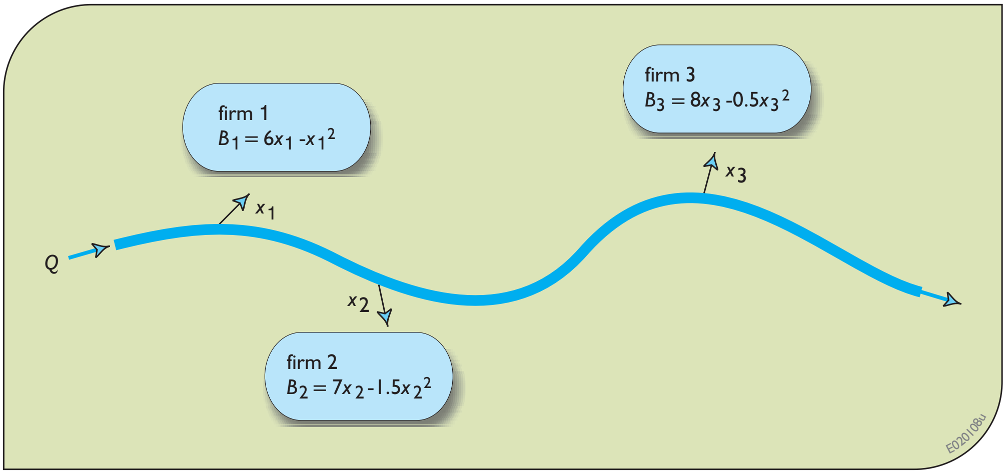



**Figure 6.9.** Water allocation to three users from a stream having a flow of *Q*.

Van den Boogaard (1998), Wen and Lee (1998) and See and Openshaw (1999). For the identification of other input–output relations see for example Haykin (1999) and Beale and Jackson (1990).

## 3. Genetic Algorithms

### 3.1. The Approach

Genetic algorithms are randomized general-purpose search techniques used for finding the best values of the parameters or decision-variables of existing models. It is not a model-building tool like genetic programming or artificial neural networks. Genetic algorithms and their variations are based on the mechanisms of natural selection (Goldberg, 1989). Unlike conventional optimization search approaches based on gradients, genetic algorithms work on a population of possible solutions, attempting to find a solution set that either maximizes or minimizes the value of a function of those solution values. This function is called the objective function. Some populations of solutions may improve the value of the objective function, others may not. The ones that improve its value play a greater role in the generation of new populations of solutions than those that do not.

Each individual solution set contains the values of all the parameters or variables whose best values are being sought. These solutions are expressed as strings of values. For example, if the values of three variables $x$, $y$ and $z$ are to be obtained, these variables are arranged into a string, $xyz$. If each variable is expressed using three digits, then the string 056004876 would represent $x = 56$, $y = 4$, and $z = 876$. These strings are called *chromosomes*. A chromosome is an array of numbers. The numbers on the chromosome are called *genes*. Pairs of chromosomes from two parents join together and produce offspring, who in turn inherit some of the genes of the parents. Altered genes may result in improved values of the objective function. These genes will tend to survive from generation to generation, while those that are inferior will tend to die.

Chromosomes are usually represented by strings of binary numbers. While much of the literature on genetic algorithms focuses on the use of binary numbers, numbers of any base may be used.

To illustrate the main features of genetic algorithms, consider the problem of finding the best allocations of water to the three water-consuming firms shown in Figure 6.9. Only integer solutions are to be considered. The maximum allocation to any single user cannot exceed 5, and the sum of all allocations cannot exceed the value of Q, say 6.

$$0 \leq x_i \leq 5 \quad \text{for } i = 1, 2, \text{ and } 3. \tag{6.7}$$

$$x_1 + x_2 + x_3 \leq 6 \tag{6.8}$$

The objective is to find the values of each allocation that maximizes the total benefits, $B(\boldsymbol{X})$.

$$\text{Maximize } B(\boldsymbol{X}) = \left(6x_1 - x_1^2\right) + \left(7x_2 - 1.5x_2^2\right) + \left(8x_3 - 0.5x_3^2\right) \tag{6.9}$$

A population of possible feasible solutions is generated randomly. A GA parameter is the size of the sample solution population – the number of solutions being considered. The best values of genetic algorithm parameters are usually determined by trial and error.

Using numbers to the base 10, a sample individual solution (chromosome) could be 312, representing the allocations $x_1 = 3$, $x_2 = 1$, and $x_3 = 2$. Another individual solution, picked at random, might be 101. These two individuals or chromosomes, each containing three genes, can pair up and have two children.

The genes of the children are determined by crossover and mutation operations. These pairing, crossover and mutation operations are random. The crossover and mutation probabilities are among the parameters of the genetic algorithm.

Suppose a crossover is to be performed on the pair of strings, 312 and 101. Crossover involves splitting the two solution strings into two parts, each string at the same place. Assume the location of the split was randomly determined to be after the first digit,

3 | 1 2

1 | 0 1

Crossover usually involves switching one part of one string with the corresponding part of the other string. After a crossover, the two new individuals are 301 and 112.

Another crossover approach is to determine for each corresponding pair of genes whether or not they will be exchanged. This would be based on some pre-set probability. For example, suppose the probability of a crossover were set at 0.30. Thus, an exchange of each corresponding pair of genes in a string or chromosome has a 30% chance of being exchanged. The result of this 'uniform' crossover involving, say, only the middle gene in the pair of strings 312 and 101 could be, say, 302 and 111. The literature on genetic algorithms describes many crossover methods for both binary as well as base-10 numbers. The interesting aspect of GA approaches is that they can be, and are, modified in many ways to suit the analyst in the search for the best solution set.

Random mutation operations apply to each gene in each string. Mutation involves changing the value of the gene being mutated. If these strings contained binary numbers, a 1 would be changed to 0, and a 0 would be changed to 1. If numbers to the base 10 are used as they are here, mutation has to be defined. Any reasonable mutation scheme can be defined. For example, suppose the mutation of a base-10 number reduces it by 1, unless the resulting number is infeasible. Hence in this example, a mutation could be defined such that if the current value of the gene being mutated (reduced) is 0, then the new number is 5. Suppose the middle digit 1 of the second new individual, 112, is randomly selected for mutation. Thus, its value changes from 1 to 0. The new string is 102. Mutation could just as well increase any number by 1 or by any other integer value. The probability of a mutation is usually much smaller than that of a crossover.

Suppose these paring, crossover and mutation operations have been carried out on numerous parent strings representing possible feasible solutions. The result is a new population of individuals (children). Each child's fitness, or objective value, can be determined. Assuming the objective function (or fitness function) is to be maximized, the higher the value the better. Adding up all the objective values associated with each child in the population, and then dividing each child's objective value by this total sum yields a fraction for each child. That fraction is the probability of that child being selected for the new population of possible solutions. The higher the objective value, the higher the probability of its becoming a parent in a new population.

In this example the objective is to maximize the total benefit derived from the allocation of water, Equation 6.9. Referring to Equation 6.9, the string 301 has a total benefit of 16.5. The string 102 has a total benefit of 19.0. The sum of these two individual benefits is 35.5. Thus the string 301 has a probability of $16.5/35.5 = 0.47$ of being selected for the new population, and the string 102 has a probability of $19/35.5 = 0.53$ of being selected. Drawing from a uniform distribution of numbers ranging from 0 to 1, if a random number is in the range 0 to 0.47, then the string 301 would be selected. If the random number exceeds 0.47, then the string 102 would be selected. Clearly in a more realistic example the new population size should be much greater than two, and indeed it typically involves hundreds of strings.

This selection or reproduction mechanism tends to transfer to the next generation the better individuals of the current generation. The higher the 'fitness' (i.e. the objective value) of an individual – in other words, the larger the relative contribution to the sum of objective function values of the entire population of individual solutions – the greater will be the chances of that

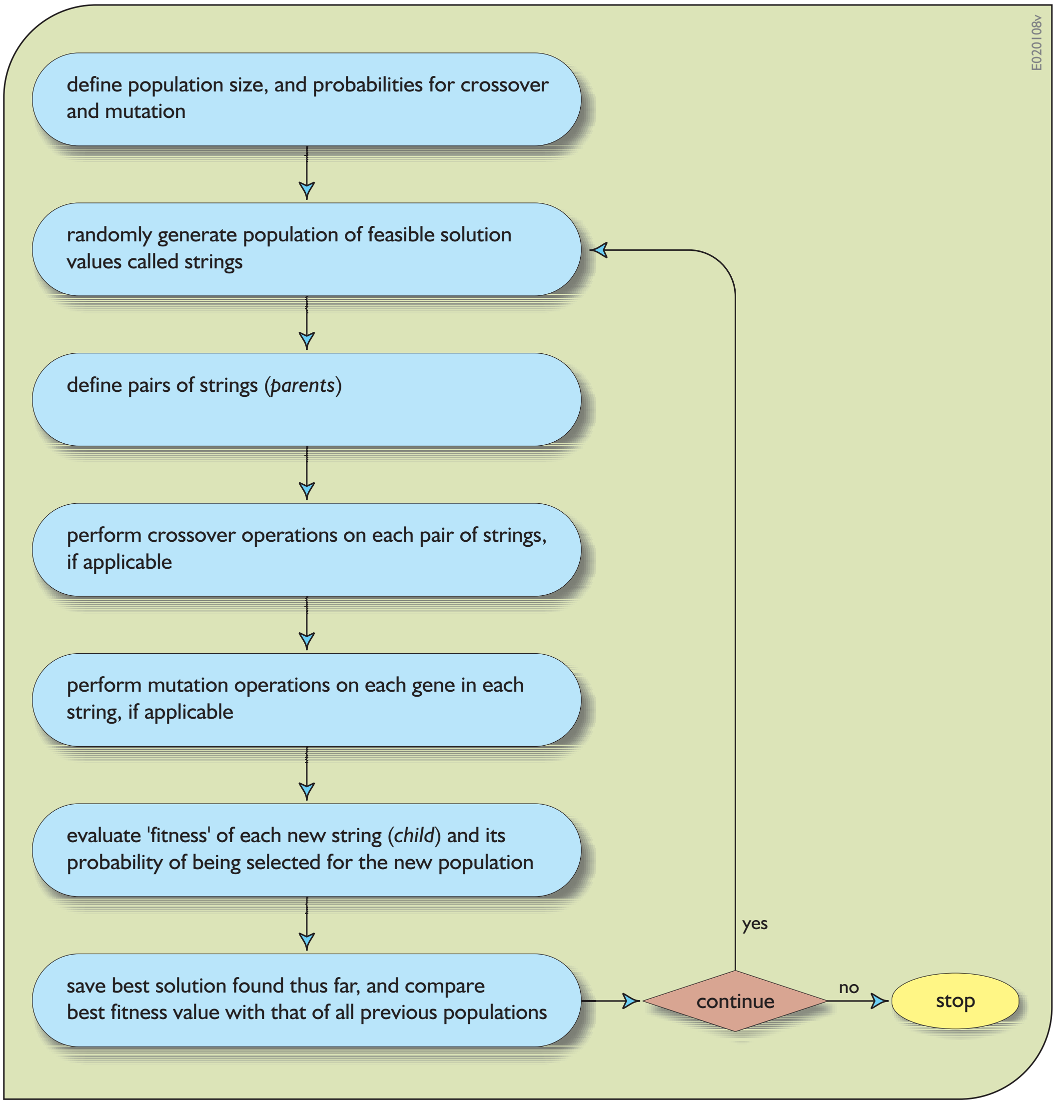



**Figure 6.10.** Flow chart of genetic algorithm procedure.

individual string of solution values being selected for the next generation.

Genetic algorithms involve numerous iterations of the operations just described. Each iteration (or generation) produces populations that tend to contain better solutions. The best solution of all populations of solutions should be saved. The genetic algorithm process can end when there is no significant change in the values of the best solution that has been found. In this search process, there is no guarantee this best solution will be the best that could be found, that is, a global optimum.

This general genetic algorithm process just described is illustrated in the flow chart in Figure 6.10.

## 3.2. Example Iterations

A few iterations with a small population of ten individual solutions for this example water-allocation problem can illustrate the basic processes of genetic algorithms. In practice, the population typically includes hundreds of individuals and the process involves hundreds of iterations. It would also likely include some procedures the modeller/programmer may think would help identify the best solution. Here we will keep the process relatively simple.

The genetic algorithm process begins with the random generation of an initial population of feasible solutions, proceeds with the paring of these solution strings, performs random crossover and mutation operations, computes the probability that each resulting child will be selected for the next population, and then randomly generates the new population. This process repeats itself with the new population and continues until there is no significant improvement in the best solution found from all past iterations.

For this example, we will

1. Randomly generate an initial population of strings of allocation variable values, ensuring that each allocation value (gene) is no less than 0 and no greater than 5. In addition, any set of allocations $x_1$, $x_2$ and $x_3$ that sums to more than 6 will be considered infeasible and discarded.
2. Pair individuals and determine if a crossover is to be performed on each pair, assuming the probability of a crossover is 50%. If a crossover is to occur, we will determine where in the string of numbers it will take place, assuming an equal probability of a crossover between any two numbers.
3. Determine if any number in the resulting individual strings is to be mutated, assuming the probability of mutation of any particular number (gene) in any string (chromosome) of numbers is 0.10. For this example, a mutation reduces the value of the number by 1, or if the original number is 0, mutation changes it to 5. After mutation, all strings of allocation values (the genes in the chromosome) that sum to more than 6 are discarded.
4. Using Equation 6.9, evaluate the 'fitness' (total benefits) associated with the allocations represented by each individual string in the population. Record the best individual string of allocation values from this and previous populations.
5. Return to Step 1 above if the change in the best solution and its objective function value is significant; Otherwise terminate the process.

These steps are performed in Table 6.3 for three iterations using a population of 10.

The best solution found so far is 222: that is, $x_1 = 2$, $x_2 = 2$, $x_3 = 2$. This process can and should continue. Once the process has converged on the best solution it can find, it may be prudent to repeat the process, but this time, change the probabilities of crossover or mutation or let mutation be an increase in the value of a number rather than a decrease. It is easy to modify the procedures used by genetic algorithms in an attempt to derive the best solution in an efficient manner.

## 4. Genetic Programming

One of the challenges in computer science is to learn how to program computers to perform a task without telling them how to do it. In other words, how can we enable computers to learn to program themselves for solving particular problems? Since the 1950s, computer scientists have tried, with varying degrees of success, to give computers the ability to learn. The name for this field of study is 'machine learning' (ML), a phrase used in 1959 by the first person to make a computer perform a serious learning task, Arthur Samuel. Originally, 'machine learning' meant the ability of computers to program themselves. That goal has, for many years, proven very difficult. As a consequence, computer scientists have pursued more modest goals. A good present-day definition of machine learning is given by Mitchell (1997), who identifies machine learning as the study of computer algorithms that improve automatically through experience.

Genetic programming (GP) aspires to do just that: to induce a population of computer programs or models (objects that turn inputs to outputs) that improve automatically as they experience the data on which they are trained (Banzhaf et al., 1998). Genetic programming is one of many machine-learning methods. Within the machine learning community, it is common to use 'genetic programming' as shorthand for any machine learning system that evolves tree structures (Koza et al., 1992).

While there is no GP today that will automatically generate a model to solve any problem, there are some examples where GP has evolved programs that are better than the best programs written by people to solve a number of difficult engineering problems. Some examples of these human-competitive GP achievements can be seen in Koza (1999), as well as in a longer list on the Internet (www.genetic-programming.com/humancompetitive.html). Since Babovic (1996) introduced the GP paradigm in the field of water engineering, a number of researchers have used the technique to analyse a variety of water management problems.

The main distinctive feature of GP is that it conducts its search for a solution to a given problem by changing model structure rather than by finding better values of model parameters or variables. There is no guarantee, however, that the resulting structure (which could be as simple as regression Equations 6.1, 6.2 or 6.3) will give us any insight into the actual workings of the system.

**Table 6.3.** Several iterations for solving the allocation problem using genetic algorithms.

| | population | crossover | | mutation | fitness | sel. prob. | cum. prob. | new pop. |
|---|---|---|---|---|---|---|---|---|
| first iteration | 230 | 23\|0 | 220 | 210 | 13.5 | 0.09 | 0.09 | 230 |
| | 220 | 22\|0 | 230 | 230 | 15.5 | 0.10 | 0.19 | 201 |
| | 021 | 021 | 021 | 021 | 15.5 | 0.10 | 0.29 | 211 |
| | 201 | 201 | 201 | 201 | 15.5 | 0.10 | 0.39 | 132 |
| | 301 | 3\|01 | 321 | 321 | (became infeasible) | | | 301 |
| | 221 | 2\|21 | 201 | 101 | 12.5 | 0.08 | 0.47 | 132 |
| | 301 | 30\|1 | 301 | 301 | 16.5 | 0.11 | 0.58 | 301 |
| | 211 | 21\|1 | 211 | 211 | 21.0 | 0.14 | 0.72 | 021 |
| | 132 | 132 | 132 | 132* | 26.5 | 0.19 | 0.19 | 230 |
| | 310 | 310 | 310 | 210 | 13.5 | 0.09 | 1.00 | 132 |
| | **total fitness** | | | | 150.0 | | | |
| second iteration | 230 | 230 | 230 | 220 | 16.0 | 0.08 | 0.08 | 221 |
| | 201 | 201 | 201 | 201 | 15.5 | 0.08 | 0.16 | 132 |
| | 211 | 21\|1 | 212 | 212* | 27.5 | 0.14 | 0.30 | 230 |
| | 132 | 13\|2 | 131 | 121 | 20.5 | 0.10 | 0.40 | 212 |
| | 301 | 301 | 301 | 301 | 16.5 | 0.08 | 0.48 | 201 |
| | 132 | 132 | 132 | 132 | 26.5 | 0.14 | 0.62 | 132 |
| | 301 | 3\|01 | 001 | 001 | 7.5 | 0.04 | 0.66 | 301 |
| | 021 | 0\|21 | 321 | 221 | 23.5 | 0.12 | 0.78 | 221 |
| | 230 | 230 | 230 | 230 | 15.5 | 0.08 | 0.86 | 212 |
| | 132 | 132 | 132 | 132 | 26.5 | 0.14 | 1.00 | 001 |
| | **total fitness** | | | | 195.5 | | | |
| third iteration | 221 | 22\|1 | 222 | 222* | 30.0 | 0.15 | 0.15 | 221 |
| | 132 | 13\|2 | 131 | 121 | 20.5 | 0.10 | 0.25 | 222 |
| | 230 | 230 | 230 | 230 | 15.5 | 0.07 | 0.32 | 230 |
| | 212 | 212 | 212 | 112 | 24.5 | 0.12 | 0.44 | 202 |
| | 201 | 20\|1 | 202 | 202 | 22.0 | 0.09 | 0.53 | 131 |
| | 132 | 13\|2 | 131 | 131 | 20.0 | 0.11 | 0.64 | 222 |
| | 301 | 301 | 301 | 201 | 15.5 | 0.08 | 0.72 | 012 |
| | 221 | 221 | 221 | 221 | 23.5 | 0.11 | 0.83 | 121 |
| | 212 | 2\|12 | 201 | 201 | 15.5 | 0.08 | 0.91 | 202 |
| | 001 | 0\|01 | 012 | 012 | 19.5 | 0.09 | 1.00 | 121 |
| | **total fitness** | | | | 206.5 | | | |

E020820e

The task of genetic programming is to find at the same time both a suitable functional form of a model and the numerical values of its parameters. To implement GP, the user must define some basic building blocks (mathematical operations and variables to be used); the algorithm then tries to build the model using the specified building blocks.

One of the successful applications of GP in automatic model-building is that of symbolic regression. Here GP searches for a mathematical regression expression in symbolic form that best produces the observed output given the associated input. To perform this task GP uses a physical symbol system divided into two sets. The first set, the so-called terminal set, contains the symbols for independent variables as well as parameter constants as appropriate. The content of this set is determined only by the nature of the problem to be solved. All the basic operators used to form a function $f(\bullet)$ are in the second set, called the functional set. For example, the second set can contain the arithmetic operators (+, −, *, /) and perhaps others such as log, sqrt, and sin as well, depending on the perceived degree of complexity of the regression.

As models have a well-defined grammar and are of variable size and shape, the structures undergoing adaptation in genetic programming are often taken from a class of parse trees. A parse tree is a node-link tree whose nodes are procedures, functions, variables and constants. The sub-trees of a node of a parse tree represent the arguments of the procedure or function of that node. Variables and functions requiring no arguments are called terminals. They are the leaves in the parse tree. Functions that take arguments are branches of a parse tree and are called functions. All terminals in one parse tree compose the terminal set, and all functions in that tree are in the functional set.

One example of a parse tree for the expression a + (b/c) can be depicted as shown in Figure 6.11.

In Figure 6.11 the variables a, b and c are leaves in the parse tree and hence belong to the terminal set. The mathematical operations + and / are functions having two arguments and, hence, are members of the functional set.

The advantage of encoding the induced GP model as a parse tree is that mutations can readily be performed by changing a node's arguments or operator function. Crossover can be achieved by replacing one or more nodes from one individual with those from another without introducing illegitimate syntax changes.

To produce new expressions (individuals) GP requires that two 'parent' expressions from the previous generation be divided and recombined into two offspring expressions. An example of this is illustrated in Figure 6.12. Having two parent expressions presented as parse trees, the crossover operation simply exchanges a branch of one parent with a branch of the other.

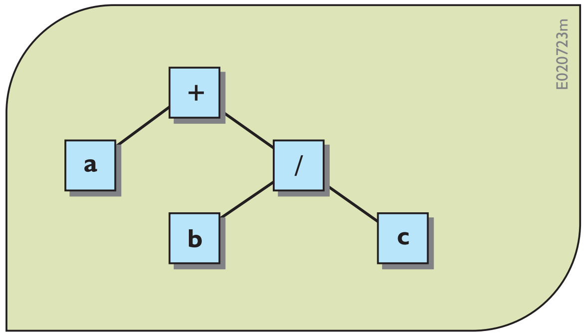



**Figure 6.11. A parse tree of the expression a + (b/c).**

The result of mutations and crossover operations is the production of two new individuals or children (Figure 6.12 lower). Each child inherits some characteristics from its parents.

This process continues until the fitness of the entire population increases and converges to find the near optimal solution set. The benefit of using the parse tree vehicle in GP can be seen in the crossover operation. This presentation guarantees that every resulting expression is grammatically and semantically correct, regardless of the crossover or mutation site, assuming that it was performed on the correct parents.

Mutation in GP corresponds to a random alteration of the individual parse tree at the branch or node level. There are several possible types of computational mutations. Some examples are:

- *Branch-mutation.* A complete sub-tree is replaced with another, possibly arbitrary sub-tree.
- *Node-mutation.* The value of a single node in the tree is replaced by another, again in an arbitrary way.
- *Constant-mutation.* Similar to node mutation, but here the constant is selected and mutated using white noise.
- *Inversion-mutation.* Inverts the order of operands in an expression in a random way.
- *Edit-mutation.* Edits the parse tree based on semantic equivalence. This mutation does not alter the function produced but just makes it shorter – for example elements like $(x + 0)$ will be replaced everywhere in the expression with $x$ only.

Mutation can affect both the parse tree structure and its information content. With mutation, therefore, the search explores a new domain. It also serves to free the search

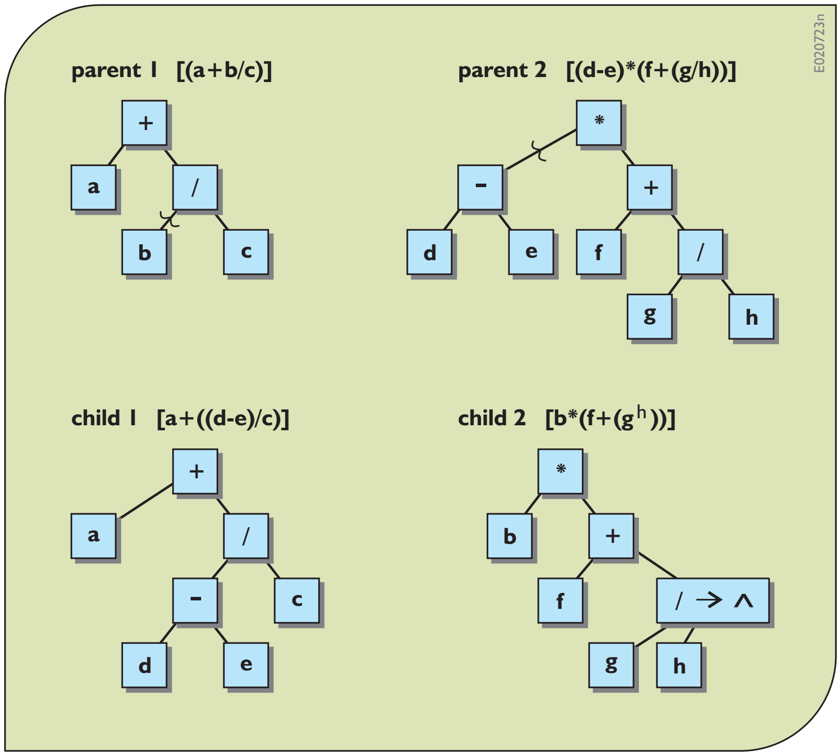



**Figure 6.12.** An illustration of a crossover operation and mutation ( / → *) operation for genetic programming.

from the possibility of being trapped in local optima. Mutation can be destructive, causing rapid degradation of relatively fit solution sets if the probability of mutation in the algorithm is set too high.

As in most evolutionary algorithms, the models that produce the best 'fit' to the data have the greatest opportunity to become parents and produce children, which is called reproduction. The better models produce the smallest errors, or differences between the calculated output and the observed output.

The basic GP procedure is as follows:

1. Generate the initial population (of size $N$) of random models (algebraic expressions). Several different initialization procedures exist for GP. Details for these methods can be seen in Babovic and Keijzer (2000).
2. Evaluate each sample model in the population and assign it a fitness value according to how well it solves the problem.
3. Create a new population of size $N$ of models by applying the following two primary operations:
   a) copy the best $M$ ($M < N$) fitted individuals – 'elite' – to the new population (asexual reproduction);
   b) create new $L$ ($L = N - M$) models by genetically recombining randomly chosen parts of two existing models (sexual reproduction).
4. Randomly apply a genetic operator mutation to the offspring and determine the fitness of each model.
5. Repeat steps 3 and 4 until a predetermined stopping criterion is reached.

The stopping criterion in Step 5 of the GP procedure is usually that either (a) a certain number of generations has been produced, or (b) a certain amount of wall-clock time has passed.

Software programs have been written to implement GP. For example GPKernel developed by Babovic and Keijzer (2000) at the Danish Hydraulic Institute (DHI) has been used in applications such as: rainfall–runoff modelling (Babovic and Abbott, 1997; Drecourt, 1999; Liong et al. 2000), sediment transport modelling, salt intrusion in estuaries and roughness estimation for a flow over a vegetation bed (Babovic and Abbott, 1997). More details about GPKernel can be seen in Aguilera (2000).

The challenge in applying genetic programming for model development is not only getting a close fit between observed and predicted outputs, given a set of input data, but also of interpreting the model that is generated to obtain additional understanding of the actual processes taking place. This 'data mining' is discussed in the next section. There are also potential problems in creating a dimensionally correct model if the input data are not dimensionless. As a consequence, many applications using GP seem to require some guidance based on a mix of both physically based and data-based approaches.

## 5. Data Mining

Data mining is the process of finding new and potentially useful knowledge from data – usually data from large databases. Data mining is also known as *knowledge discovery* (KD). It is the non-trivial extraction of implicit, previously unknown, information from data. It uses machine learning and statistical and visualization techniques to discover and present knowledge in a comprehensible form.

Data mining is a relatively new technology that has the potential to help us learn more from our data. Data mining tools can scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Data mining techniques can be implemented on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought online. When implemented on high performance client/server or parallel processing computers, data mining tools can analyse massive databases to seek answers to questions.

### 5.1. Data Mining Methods

Data mining is supported by three technologies (Babovic, 1998):

- monitoring and data collection
- multiprocessor computers
- data mining algorithms

The core components of data mining technology have been under development for decades in research areas such as statistics, artificial intelligence and machine learning. The current level of development of these techniques, coupled with high-performance relational database engines and broad data-integration efforts, make these technologies practical for current data environments.

Dynamic data access is critical for data navigation applications, and the ability to store, manage and analyse large databases is critical to data mining. Given databases of sufficient size and quality, data mining technology can provide:

- *Automated prediction of trends and behaviours.* Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data and much more quickly. A typical example of a predictive problem is identifying segments of an ecosystem likely to respond similarly to given events.
- *Automated discovery of previously unknown patterns.* Data mining tools sweep through databases and identify previously hidden patterns. When data mining tools are implemented on high performance parallel processing systems, they can analyse large databases very quickly – often in minutes. Faster processing means that users can automatically experiment with more models to understand complex data. Larger databases, in turn, yield improved predictions.

Some of the techniques commonly used in data mining are:

- *Artificial neural networks*, which are nonlinear predictive models that learn through training and resemble biological neural networks in structure, as discussed earlier.
- *Classification and regression trees* (CARTs), which is used for classification of a dataset. It provides a set of rules that one can apply to a new (unclassified) dataset to predict which records will have a given outcome. It segments a dataset by creating two-way splits.
- *Chi square automatic interaction detection* (CHAID), which is a decision tree technique used for classification of a dataset. It provides a set of rules that one can apply to a new (unclassified) dataset to predict which records will have a given outcome. It segments a dataset by using chi square tests to create multi-way splits. It requires more data preparation than CART.
- *Classification*, which is the process of dividing a dataset into mutually exclusive groups, such that the members of

each group are as 'close' as possible to one another, and different groups are as 'far' as possible from one another, where distance is measured with respect to specific variable(s) one is trying to predict.

- *Clustering*, which is similar to classification, except that distance is measured with respect to all available variables.
- *Data navigation*, which is the process of viewing different dimensions, slices and levels of detail of a multidimensional database. See OLAP, defined below.
- *Data visualization*, which is the visual interpretation of complex relationships in multidimensional data.
- *Decision trees*, which are tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include classification and regression trees (CARTs) and chi square automatic interaction detection (CHAID).
- *Genetic algorithms and genetic programming*, which search techniques that use processes such as genetic combination, mutation and natural selection in a design based on the concepts of evolution, as discussed earlier.
- *Nearest-neighbour method*, which is a technique that classifies each record in a dataset on the basis of a combination of the classes of the *k* record(s) most similar to it in a historical dataset. It is sometimes called the *k*-nearest-neighbour technique.
- *Online analytical processing* (OLAP), which refers to array-oriented database applications that allow users to view, navigate through, manipulate and analyse multidimensional databases.
- *Rule induction*, which is the method of extraction of useful if–then rules from data based on statistical significance.

Many of these technologies have been in use for more than a decade in specialized analysis tools that work with relatively small volumes of data. These capabilities are now evolving to integrate directly with standard databases and online application program platforms.

## 6. Conclusions

Most computer-based models used for water resources planning and management are physical, mechanistic or process-based models. Builders of such models attempt to incorporate the important physical, biological, chemical, geomorphological, hydrological and other types of interactions among all system components, as appropriate for the problem being addressed and system being modelled. This is done in order to be able to predict possible economic, ecologic, environmental or social impacts that might result from the implementation of some plan or policy. These types of models almost always contain parameters. These need values, and the values of the parameters affect the accuracy of the impact predictions.

This chapter has introduced some data-based methods of modelling. These have included two evolutionary search approaches: genetic algorithms (GA) for estimating the parameter values, and genetic programming for finding models that replicate the real system. In some situations, these biologically motivated search methods, which are independent of the particular model being calibrated, provide the most practical way for model parameter calibrations to be accomplished. These same search methods, sometimes coupled to physically based simulation models, can be used to obtain the values of the unknown decision-variables as well.

While physically based or process-based models are appealing for those who wish to better understand these natural processes, they clearly remain approximations of reality. In some water resources problem applications, the complexities of the real system are considerably greater than the complexities of the models built to simulate them. Hence, it should not be surprising that in some cases statistical or data-based models, which convert input variable values to output variable values in ways that have nothing to do with what happens in reality, may produce more accurate results than physically-based models. This chapter has briefly introduced two such types of data-based model: genetic programming (GP) models and artificial neural networks (ANN). When they work, they often produce results faster than their physical counterparts and as accurately or more so, but only within the range of values observed in the data used to build these models.

Data-driven modelling methods are increasingly being developed and used to gain information from data. They are especially useful when the data sets are large, and where it becomes impractical for any human to sort through it to obtain further insights into the processes that produced the data. Many data-rich problems can be solved by using novel data-driven modelling together with other techniques. Data mining methods are also being increasingly used to gain greater understanding and knowledge from large data sets. Some approaches to data mining are listed in this chapter. Water resources modellers are unlikely to be involved in the development of such data mining techniques, so fortunately, as is the case with GA, GP and ANN methods, many are available on the Internet. Applications of such methods to groundwater modelling, sedimentation processes along coasts and in harbours, rainfall runoff prediction, reservoir operation, data classification, and predicting surge water levels for navigation represent only a small sample of what can be found in the current literature. Some of this literature is cited in the next section.

## 7. References

ABRAHART, R.J.; KNEALE, P.E. and SEE, L.M. (eds.). 2004. *Neural networks for hydrological modeling*. Leiden, the Netherlands, AA Balkema.

AGUILERA, D.R. 2000. *Genetic Programming with GPKERNEL*. Unpublished, DHI, Copenhagen.

BABOVIC, V. 1996. *Emergence, evolution, intelligence: hydroinformatics*.

BABOVIC, V. 1998. A data mining approach to time series modelling and forecasting. In: V. Babovic and L.C. Larsen (eds.), *Proceedings of the Third International Conference on Hydroinformatics (Hydroinformatics '98)*, Rotterdam, Balkema, pp. 847–56.

BABOVIC, V. and ABBOTT, M.B. 1997. The evolution of equations from hydraulic data. Part II: applications. *Journal of Hydraulic Research*, Vol. 35, No. 3, pp. 411–27.

BABOVIC, V. and BOJKOV, V.H. 2001. *D2K technical report D2K TR 0401-1*. Copenhagen, DHI, April.

BABOVIC, V. and KEIJZER, M. 2000. Genetic programming as a model induction engine. *Journal of Hydroinformatics*, Vol. 2, No. 1, pp. 35–60.

BANZHAF, W.; NORDIN, P.; KELLER, R.E. and FRANNKONE, F.D. 1998. *Genetic programming: an introduction: on the automatic evolution of computer programs and its applications*. San Francisco, Calif., Morgan Kaufmann Publishers, Inc. and dpunkt – Verlag fur Digitale Technologie GmbH.

BEALE, R. and JACKSON, T. 1990. *Neural computing: an introduction*. Bristol, IOP.

CLAIR, T.A. and EHRMAN, J.M. 1998. Using neural networks to assess the influence of changing seasonal climates in modifying discharge, dissolved organic carbon, and nitrogen export in eastern Canadian rivers. *Water Resources Research*, Vol. 34, No. 3, pp. 447–55.

CYBENKO, G. 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, Vol. 2, pp. 303–14.

DRECOURT, J.-P. 1999. *Application of neural networks and genetic programming to rainfall–runoff modelling*, D2K Technical Report 0699-1. Copenhagen, DHI.

GOLDBERG, D.E. 1989. *Genetic algorithms in search, optimization, and machine learning*. Reading, Addison-Westley.

HAGAN, M.T.; DEMUTH, H.B. and BEALE, M. 1996. *Neural network design*. Boston, Mass., PWS.

HAYKIN, S. 1999. *Neural networks: a comprehensive foundation*, 2nd edn. Upper Saddle River, N.J., Prentice-Hall.

HERTZ, J.; KROGH, A. and PALMER, R.G. 1991. *Introduction to the theory of neural computation*. Reading, Mass., Addison Wesley.

HORNIK, K.; STINCHCOMBE, M. and WHITE, H. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks*, No. 2, pp. 359–66.

HSIEH, W.W. and TANG, B. 1998. Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society*, Vol. 79, pp. 1855–70.

KOZA, J.R. 1992. *Genetic programming: on the programming of computers by means of natural selection (complex adaptive systems)*. Cambridge, Mass., MIT Press.

KOZA, J.R.; BENNETT, F.H. III; ANDRE, D. and KEANE, M.A. 1999. *Genetic programming III: Darwinian invention and problem solving*. San Francisco, Calif., Morgan Kaufmann.

LANGE, N.T.G. 1998. Advantages of unit hydrograph derivation by neural networks. In: V. Babovic and L.C. Larsen (eds.), *Hydroinformatics '98*, Rotterdam, Balkema, pp. 783–89.

LIONG, S.-Y.; KHU, S.T.; BABOVIC, V.; HAVNOE, K.; CHAN, W.T. and PHOON, K.K. 2000. *Construction of non-linear models in water resources with evolutionary computation*. Singapore, National University of Singapore.

MINNS, A.W. 1996. Extended rainfall–runoff modelling using artificial neural networks. In: A. Müller (ed.), *Hydroinformatics '96*, Rotterdam, Balkema, pp. 207–13.

MINNS, A.W. and HALL, M.J. 1996. Artificial neural networks as rainfall–runoff models. *Hydrological Sciences Journal*, Vol. 41, No. 3, pp. 399–417.

MITCHELL, T.M. 1997. *Machine learning*. New York, McGraw-Hill.

PRICE, R.K.; SAMEDOV, J.N. and SOLOMATINE, D.P. 1998. An artificial neural network model of a generalised channel network. In: V. Babovic and L.C. Larsen (eds.), *Hydroinformatics '98*, Rotterdam, Balkema, pp. 813–18.

PRINCIPE, J.C.; EULIANO, N.R. and LEFEBVRE, W.C. 2000. *Neural and adaptive systems: fundamentals through simulation*. New York, John Wiley.

PROAÑO, C.O.; VERWEY, A.; VAN DEN BOOGAARD, H.F.P. and MINNS, A.W. 1998. Emulation of a sewerage system computational model for the statistical processing of large number of simulations. In: V. Babovic and L.C. Larsen (eds.), *Hydroinformatics '98*, Vol. 2, Rotterdam, Balkema, pp. 1145–52.

RECKNAGEL, F.; FRENCH, M.; HARKONEN P. and YANUNAKA, K.I. 1997. Artificial neural network approach for modelling and prediction of algal blooms. *Ecological Modelling*, No. 96, pp. 11–28.

SANCHEZ, L.; ARROYO, V.; GARCIA, J.; KOEV, K. and REVILLA, J. 1998. Use of neural networks in design of coastal sewage systems. *Journal of Hydraulic Engineering*, Vol. 124, No. 5, pp. 457–64.

SCARDI, M. 1996. Artificial neural networks as empirical models for estimating phytoplankton production. *Marine Ecology Progress Series*, No. 139, pp. 289–99.

SEE, L. and OPENSHAW, S. 1999. Applying soft computing approaches to river level forecasting. *Hydrological Sciences Journal*, Vol. 44, No. 5, pp. 763–78.

SHEN, Y.; SOLOMATINE, D.P. and VAN DEN BOOGAARD, H.F.P. 1998. Improving performance of chlorophyl concentration time series simulation with artificial neural networks. *Annual Journal of Hydraulic Engineering, JSCE*, No. 42, pp. 751–6.

SMITH, M. 1993. *Neural networks for statistical modelling*. New York, Van Nostrand Reinhold.

STELLING, G.S. 2000. A numerical method for inundation simulations. In: Y.N. Yoon, B.H. Jun, B.H. Seoh and G.W. Choi (eds.), *Proceedings of the 4th International Conference on Hydro-Science and Engineering*, Seoul, 26–26 September 2000. Seoul, Korea Water Resources Association.

VAN DEN BOOGAARD, H.F.P.; GAUTAM, D.K. and MYNETT, A.E. 1998. Auto-regressive neural networks for the modelling of time series. In: V. Babovic and L.C. Larsen (eds.), *Hydroinformatics '98*, Rotterdam, Balkema, pp. 741–8.

VAN DEN BOOGAARD, H.F.P.; TEN BRUMMELHUIS, P.G.J. and MYNETT, A.E. 2000. *On-line data assimilation in auto-regressive neural networks*. Proceedings of the Hydroinformatics 2000 Conference, The University of Engineering, Iowa, USA, July 23–27, 2000. Iowa City, University of Iowa.

VAN GENT, M.R.A. and VAN DEN BOOGAARD, H.F.P. 1998. Neural network modelling of forces on vertical structures. In: *Proceedings of the 27th International Conference on Coastal Engineering*, pp. 2096–109. Reston, Va., ASCE press.

WEN, C.-G. and LEE, C.-S. 1998. A neural network approach to multiobjective optimization for water quality management in a river basin. *Water Resources Research*, Vol. 34, No. 3, pp. 427–36.

WÜST, J.C. 1995. Current prediction for shipping guidance. In: B. Kappen and S. Gielen (eds.). *Neural networks: artificial intelligence and industrial applications*. Proceedings of the 3rd Annual SNN Symposium on Neural Networks, Nijmegen, The Netherlands, 14–15 September 1995. London: Springer-Verlag, pp. 366–73.

## Additional References (Further Reading)

MCKINNEY, D.C. and LIN, M.D. 1994. Genetic algorithm solution of groundwater management models. *Water Resources Research*, Vol. 30, No. 6, pp. 1897–906.

OLIVERA, R. and LOUCKS, D.P. 1997. Operating rules for multi-reservoir operation. *Water Resources Research*, Vol. 33, No. 4, pp. 839–52.

SOLOMATINE, D.P. and AVILA TORRES, L.A. 1996. Neural network application of a hydrodynamic model in optimizing reservoir operation. In: A. Müller (ed.), *Hydroinformatics '96*, Rotterdam, Balkema, pp. 201–06.

# 7. Concepts in Probability, Statistics and Stochastic Modelling

1. Introduction 169

2. Probability Concepts and Methods 170
   - 2.1. Random Variables and Distributions 170
   - 2.2. Expectation 173
   - 2.3. Quantiles, Moments and Their Estimators 173
   - 2.4. L-Moments and Their Estimators 176

3. Distributions of Random Events 179
   - 3.1. Parameter Estimation 179
   - 3.2. Model Adequacy 182
   - 3.3. Normal and Lognormal Distributions 186
   - 3.4. Gamma Distributions 187
   - 3.5. Log-Pearson Type 3 Distribution 189
   - 3.6. Gumbel and GEV Distributions 190
   - 3.7. L-Moment Diagrams 192

4. Analysis of Censored Data 193

5. Regionalization and Index-Flood Method 195

6. Partial Duration Series 196

7. Stochastic Processes and Time Series 197
   - 7.1. Describing Stochastic Processes 198
   - 7.2. Markov Processes and Markov Chains 198
   - 7.3. Properties of Time-Series Statistics 201

8. Synthetic Streamflow Generation 203
   - 8.1. Introduction 203
   - 8.2. Streamflow Generation Models 205
   - 8.3. A Simple Autoregressive Model 206
   - 8.4. Reproducing the Marginal Distribution 208
   - 8.5. Multivariate Models 209
   - 8.6. Multi-Season, Multi-Site Models 211
     - 8.6.1. Disaggregation Models 211
     - 8.6.2. Aggregation Models 213

9. Stochastic Simulation 214
   - 9.1. Generating Random Variables 214
   - 9.2. River Basin Simulation 215
   - 9.3. The Simulation Model 216
   - 9.4. Simulation of the Basin 216
   - 9.5. Interpreting Simulation Output 217

10. Conclusions 223

11. References 223

# 7 Concepts in Probability, Statistics and Stochastic Modelling

Events that cannot be predicted precisely are often called random. Many if not most of the inputs to, and processes that occur in, water resources systems are to some extent random. Hence, so too are the outputs or predicted impacts, and even people's reactions to those outputs or impacts. To ignore this randomness or uncertainty is to ignore reality. This chapter introduces some of the commonly used tools for dealing with uncertainty in water resources planning and management. Subsequent chapters illustrate how these tools are used in various types of optimization, simulation and statistical models for impact prediction and evaluation.

## 1. Introduction

Uncertainty is always present when planning, developing, managing and operating water resources systems. It arises because many factors that affect the performance of water resources systems are not and cannot be known with certainty when a system is planned, designed, built, managed and operated. The success and performance of each component of a system often depends on future meteorological, demographic, economic, social, technical, and political conditions, all of which may influence future benefits, costs, environmental impacts, and social acceptability. Uncertainty also arises due to the stochastic nature of meteorological processes such as evaporation, rainfall and temperature. Similarly, future populations of towns and cities, per capita water-usage rates, irrigation patterns and priorities for water uses, all of which affect water demand, are never known with certainty.

There are many ways to deal with uncertainty. One, and perhaps the simplest, approach is to replace each uncertain quantity either by its average (i.e., its mean or expected value), its median, or by some critical (e.g., 'worst-case') value, and then proceed with a deterministic approach. Use of *expected* or *median values* of uncertain quantities may be adequate if the uncertainty or variation in a quantity is reasonably small and does not critically affect the performance of the system. If expected or median values of uncertain parameters or variables are used in a deterministic model, the planner can then assess the importance of uncertainty by means of sensitivity analysis, as is discussed later in this and the two subsequent chapters.

Replacement of uncertain quantities by either expected, median or worst-case values can grossly affect the evaluation of project performance when important parameters are highly variable. To illustrate these issues, consider the evaluation of the recreation potential of a reservoir. Table 7.1 shows that the elevation of the water surface varies over time depending on the inflow and demand for water. The table indicates the pool levels and their associated probabilities as well as the expected use of the recreation facility with different pool levels.

The average pool level $\bar{L}$ is simply the sum of each possible pool level times its probability, or

$$\bar{L} = 10(0.10) + 20(0.25) + 30(0.30) + 40(0.25) + 50(0.10) = 30 \tag{7.1}$$

This pool level corresponds to 100 visitor-days per day:

$$VD(\bar{L}) = 100 \text{ visitor-days per day} \tag{7.2}$$

A worst-case analysis might select a pool level of ten as a critical value, yielding an estimate of system performance equal to 100 visitor-days per day:

$$VD(L_{\text{low}}) = VD(10) = 25 \text{ visitor-days per day} \tag{7.3}$$

| possible pool levels | probability of each level | recreation potential in visitor-days per day for reservoir with different pool levels |
|---|---|---|
| 10 | 0.10 | 25 |
| 20 | 0.25 | 75 |
| 30 | 0.30 | 100 |
| 40 | 0.25 | 80 |
| 50 | 0.10 | 70 |

E021101a

**Table 7.1.** Data for determining reservoir recreation potential.

Neither of these values is a good approximation of the average visitation rate, that is

$$\begin{aligned}\overline{VD} &= 0.10\ VD(10) + 0.25\ VD(20) + 0.30\ VD(30) \\ &\quad + 0.25\ VD(40) + 0.10\ VD(50) \\ &= 0.10(25) + 0.25(75) + 0.30(100) + 0.25(80) \\ &\quad + 0.10(70) \qquad (7.4) \\ &= 78.25 \text{ visitor-days per day}\end{aligned}$$

Clearly, the average visitation rate, $\overline{VD} = 78.25$, the visitation rate corresponding to the average pool level $VD(\overline{L}) = 100$, and the worst-case assessment $VD(L_{low}) = 25$, are very different.

Using only average values in a complex model can produce a poor representation of both the average performance and the possible performance range. When important quantities are uncertain, a comprehensive analysis requires an evaluation of both the expected performance of a project and the risk and possible magnitude of project failures in a physical, economic, ecological and/or social sense.

This chapter reviews many of the methods of probability and statistics that are useful in water resources planning and management. Section 2 is a condensed summary of the important concepts and methods of probability and statistics. These concepts are applied in this and subsequent chapters of this book. Section 3 presents several probability distributions that are often used to model or describe the distribution of uncertain quantities. The section also discusses methods for fitting these distributions using historical information, and methods of assessing whether the distributions are adequate representations of the data. Sections 4, 5 and 6 expand upon the use of these mathematical models, and discuss alternative parameter estimation methods.

Section 7 presents the basic ideas and concepts of the stochastic processes or time series. These are used to model streamflows, rainfall, temperature or other phenomena whose values change with time. The section contains a description of Markov chains, a special type of stochastic process used in many stochastic optimization and simulation models. Section 8 illustrates how synthetic flows and other time-series inputs can be generated for stochastic simulations. Stochastic simulation is introduced with an example in Section 9.

Many topics receive only brief treatment in this introductory chapter. Additional information can be found in applied statistical texts or book chapters such as Benjamin and Cornell (1970), Haan (1977), Kite (1988), Stedinger et al. (1993), Kottegoda and Rosso (1997), and Ayyub and McCuen (2002).

## 2. Probability Concepts and Methods

This section introduces the basic concepts and definitions used in analyses involving probability and statistics. These concepts are used throughout this chapter and later chapters in the book.

### 2.1. Random Variables and Distributions

The basic concept in probability theory is that of the *random variable*. By definition, the value of a random variable cannot be predicted with certainty. It depends, at least in part, on the outcome of a chance event. Examples are: (1) the number of years until the flood stage of a river washes away a small bridge; (2) the number of times during a reservoir's life that the level of the pool will drop below a specified level; (3) the rainfall depth next month; and (4) next year's maximum flow at a gauge site on an unregulated stream. The values of all of these random events or variables are not knowable before the event has occurred. Probability can be used to describe the likelihood that these random variables will equal specific values or be within a given range of specific values.

The first two examples illustrate *discrete random variables*, random variables that take on values that are discrete (such as positive integers). The second two examples illustrate *continuous random variables*. Continuous random variables take on any values within a specified range of values. A property of all continuous random variables is that the probability that the value of any of those random variables will equal some specific number – any specific number – is always zero. For example, the probability that the total rainfall depth in a month will be exactly 5.0 cm is zero, while the probability that the total rainfall will lie between 4 and 6 cm could be nonzero. Some random variables are combinations of continuous and discrete random variables.

Let $X$ denote a random variable and $x$ a possible value of that random variable $X$. Random variables are generally denoted by capital letters, and particular values they take on by lowercase letters. For any real-valued random variable $X$, its *cumulative distribution function* $F_X(x)$, often denoted as just the cdf, equals the probability that the value of $X$ is less than or equal to a specific value or threshold $x$:

$$F_X(x) = \Pr[X \le x] \tag{7.5}$$

This cumulative distribution function $F_X(x)$ is a non-decreasing function of $x$ because

$$\Pr[X \le x] \le \Pr[X \le x + \delta] \quad \text{for} \quad \delta > 0 \tag{7.6}$$

In addition,

$$\lim_{x \to +\infty} F_X(x) = 1 \tag{7.7}$$

and

$$\lim_{x \to -\infty} F_X(x) = 0 \tag{7.8}$$

The first limit equals 1 because the probability that $X$ takes on some value less than infinity must be unity; the second limit is zero because the probability that $X$ takes on no value must be zero.

The left half of Figure 7.1 illustrates the cumulative distribution function (upper) and its derivative, the probability density function, $f_X(x)$, (lower) of a continuous random variable $X$.

If $X$ is a real-valued discrete random variable that takes on specific values $x_1, x_2, \ldots$, then the *probability mass function* $p_X(x_i)$ is the probability $X$ takes on the value $x_i$.

$$p_X(x_i) = \Pr[X = x_i] \tag{7.9}$$

The value of the cumulative distribution function $F_X(x)$ for a discrete random variable is the sum of the probabilities of all $x_i$ that are less than or equal to $x$.

$$F_X(x) = \sum_{x_i \le x} p_X(x_i) \tag{7.10}$$

The right half of Figure 7.1 illustrates the cumulative distribution function (upper) and the probability mass function $p_X(x_i)$ (lower) of a discrete random variable.

The *probability density function* $f_X(x)$ (lower left plot in Figure 7.1) for a continuous random variable $X$ is the analogue of the probability mass function (lower right plot in Figure 7.1) of a discrete random variable $X$. The probability density function, often called the pdf, is the derivative of the cumulative distribution function so that:

$$f_X(x) = \frac{dF_X(x)}{dx} \ge 0 \tag{7.11}$$

It is necessary to have

$$\int_{-\infty}^{+\infty} f_X(x) = 1 \tag{7.12}$$

Equation 7.12 indicates that the area under the probability density function is 1. If $a$ and $b$ are any two constants, the cumulative distribution function or the density function may be used to determine the probability that $X$ is greater than $a$ and less than or equal to $b$ where

$$\Pr[a < X \le b] = F_X(b) - F_X(a) = \int_a^b f_X(x)dx \tag{7.13}$$

The area under a probability density function specifies the relative frequency with which the value of a continuous random variable falls within any specified range of values, that is, any interval along the horizontal axis.

Life is seldomly so simple that only a single quantity is uncertain. Thus, the joint probability distribution of two or more random variables can also be defined. If $X$ and $Y$ are two continuous real-valued random variables, their joint cumulative distribution function is:

$$\begin{aligned} F_{XY}(x, y) &= \Pr[X \le x \text{ and } Y \le y] \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v)du\, dv \end{aligned} \tag{7.14}$$

If two random variables are discrete, then

$$F_{XY}(x, y) = \sum_{x_i \le x} \sum_{y_i \le y} p_{XY}(x_i, y_i) \tag{7.15}$$

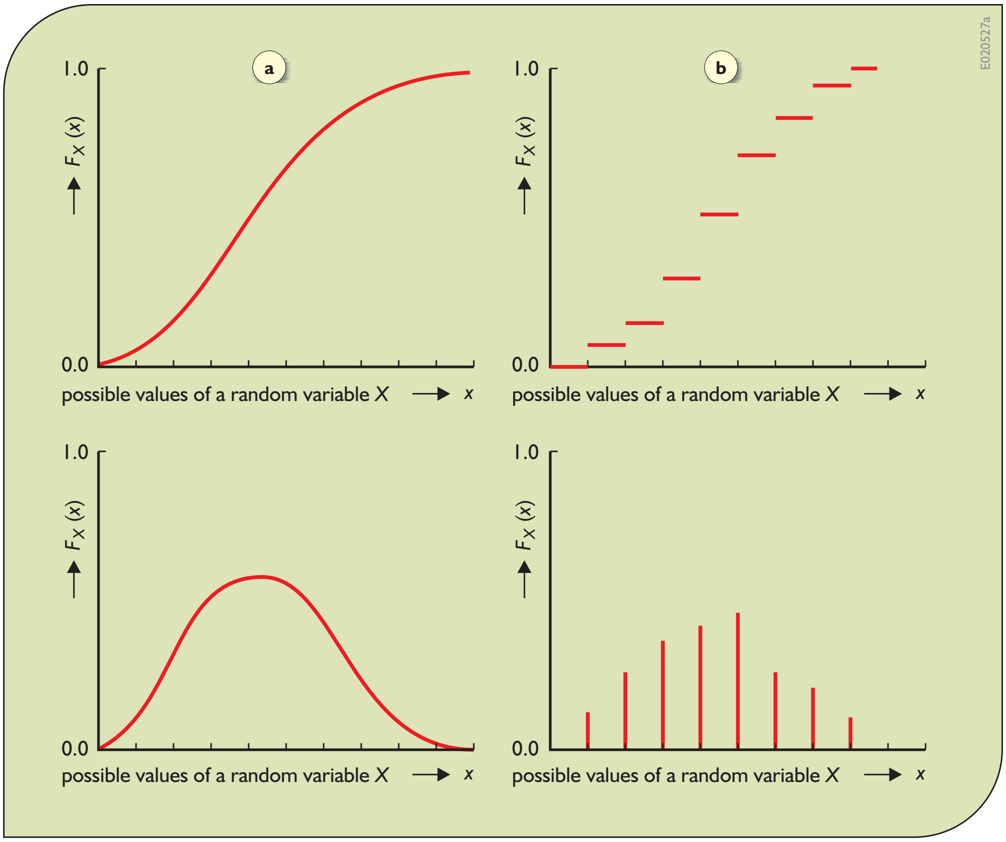



**Figure 7.1.** Cumulative distribution and probability density or mass functions of random variables: (a) continuous distributions; (b) discrete distributions.

where the joint probability mass function is:

$$p_{XY}(x_i, y_i) = \Pr[X = x_i \text{ and } Y = y_i] \tag{7.16}$$

If $X$ and $Y$ are two random variables, and the distribution of $X$ is not influenced by the value taken by $Y$, and vice versa, then the two random variables are said to be *independent*. For two independent random variables $X$ and $Y$, the joint probability that the random variable $X$ will be between values $a$ and $b$ and that the random variable $Y$ will be between values $c$ and $d$ is simply the product of those separate probabilities.

$$\begin{aligned}&\Pr[a \le X \le b \quad \text{and} \quad c \le Y \le d]\\ &\quad = \Pr[a \le X \le b] \times \Pr[c \le Y \le d]\end{aligned} \tag{7.17}$$

This applies for any values $a$, $b$, $c$, and $d$. As a result,

$$F_{XY}(x, y) = F_X(x)F_Y(y) \tag{7.18}$$

which implies for continuous random variables that

$$f_{XY}(x, y) = f_X(x)f_Y(y) \tag{7.19}$$

and for discrete random variables that

$$p_{XY}(x, y) = p_X(x)p_Y(y) \tag{7.20}$$

Other useful concepts are those of the *marginal* and *conditional distributions*. If $X$ and $Y$ are two random variables whose joint cumulative distribution function $F_{XY}(x, y)$ has been specified, then $F_X(x)$, the marginal cumulative distribution of $X$, is just the cumulative distribution of $X$ ignoring $Y$. The marginal cumulative distribution function of $X$ equals

$$F_X(x) = \Pr[X \le x] = \lim_{y \to \infty} F_{XY}(x, y) \tag{7.21}$$

where the limit is equivalent to letting $Y$ take on any value. If $X$ and $Y$ are continuous random variables, the marginal density of $X$ can be computed from

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)\mathrm{d}y \tag{7.22}$$

The conditional cumulative distribution function is the cumulative distribution function for $X$ given that $Y$ has taken a particular value $y$. Thus the value of $Y$ may have been observed and one is interested in the resulting conditional distribution for the so far unobserved value of $X$. The conditional cumulative distribution function for continuous random variables is given by

$$F_{X|Y}(x \mid y) = \Pr[X \leq x \mid Y = y] = \frac{\int_{-\infty}^{x} f_{XY}(s, y)\,\mathrm{d}s}{f_Y(y)} \tag{7.23}$$

where the conditional density function is

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)} \tag{7.24}$$

For discrete random variables, the probability of observing $X = x$, given that $Y = y$ equals

$$p_{X|Y}(x \mid y) = \frac{p_{XY}(x, y)}{p_Y(y)} \tag{7.25}$$

These results can be extended to more than two random variables. Kottegoda and Rosso (1997) provide more detail.

## 2.2. Expectation

Knowledge of the probability density function of a continuous random variable, or of the probability mass function of a discrete random variable, allows one to calculate the expected value of any function of the random variable. Such an expectation may represent the average rainfall depth, average temperature, average demand shortfall or expected economic benefits from system operation. If $g$ is a real-valued function of a continuous random variable $X$, the expected value of $g(X)$ is:

$$\mathrm{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x)\,\mathrm{d}x \tag{7.26}$$

whereas for a discrete random variable

$$\mathrm{E}[g(X)] = \sum_i g(x_i) p_X(x_i) \tag{7.27}$$

The *expectation operator*, $\mathrm{E}[\cdot]$, has several important properties. In particular, the expectation of a linear function of $X$ is a linear function of the expectation of $X$. Thus, if $a$ and $b$ are two non-random constants,

$$\mathrm{E}[a + bX] = a + b\mathrm{E}[X] \tag{7.28}$$

The expectation of a function of two random variables is given by

$$\mathrm{E}[g(X, Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x, y) f_{XY}(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

or

$$\mathrm{E}[g(X, Y)] = \sum_i \sum_j g(x_i, y_i) p_{XY}(x_i, y_i) \tag{7.29}$$

If $X$ and $Y$ are independent, the expectation of the product of a function $g(\cdot)$ of $X$ and a function $h(\cdot)$ of $Y$ is the product of the expectations:

$$\mathrm{E}[g(X)\,h(Y)] = \mathrm{E}[g(X)]\,\mathrm{E}[h(Y)] \tag{7.30}$$

This follows from substitution of Equations 7.19 and 7.20 into Equation 7.29.

## 2.3. Quantiles, Moments and Their Estimators

While the cumulative distribution function provides a complete specification of the properties of a random variable, it is useful to use simpler and more easily understood measures of the central tendency and range of values that a random variable may assume. Perhaps the simplest approach to describing the distribution of a random variable is to report the value of several quantiles. The $p$th quantile of a random variable $X$ is the smallest value $x_p$ such that $X$ has a probability $p$ of assuming a value equal to or less than $x_p$:

$$\Pr[X < x_p] \leq p \leq \Pr[X \leq x_p] \tag{7.31}$$

Equation 7.31 is written to insist if at some value $x_p$, the cumulative probability function jumps from less than $p$ to more than $p$, then that value $x_p$ will be defined as the $p$th quantile even though $F_X(x_p) \neq p$. If $X$ is a continuous random variable, then in the region where $f_X(x) > 0$, the quantiles are uniquely defined and are obtained by solution of

$$F_X(x_p) = p \tag{7.32}$$

Frequently reported quantiles are the *median* $x_{0.50}$ and the *lower* and *upper quartiles* $x_{0.25}$ and $x_{0.75}$. The median describes the location or central tendency of the distribution of $X$ because the random variable is, in the continuous case, equally likely to be above as below that value. The interquartile range $[x_{0.25}, x_{0.75}]$ provides an easily understood description of the range of values that the random variable might assume. The $p$th quantile is also the 100 $p$ percentile.

In a given application – particularly when safety is of concern – it may be appropriate to use other quantiles. In floodplain management and the design of flood control structures, the 100-year flood $x_{0.99}$ is a commonly selected design value. In water quality management, a river's minimum seven-day-average low flow expected once in ten years is commonly used in the United States as the critical planning value: Here the one-in-ten year value is the $10^{th}$ percentile of the distribution of the annual minima of the seven-day average flows.

The natural sample estimate of the median $x_{0.50}$ is the median of the sample. In a sample of size $n$ where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ are the observations ordered by magnitude, and for a non-negative integer $k$ such that $n = 2k$ (even) or $n = 2k + 1$ (odd), the sample estimate of the median is

$$\hat{x}_{0.50} = \begin{cases} x_{(k+1)} & \text{for} \quad n = 2k + 1 \\ \frac{1}{2}\left[x_{(k)} + x_{(k+1)}\right] & \text{for} \quad n = 2k \end{cases} \tag{7.33}$$

Sample estimates of other quantiles may be obtained by using $x_{(i)}$ as an estimate of $x_q$ for $q = i/(n + 1)$ and then interpolating between observations to obtain $\hat{x}_p$ for the desired $p$. This only works for $1/(n + 1) \leq p \leq n/(n+1)$ and can yield rather poor estimates of $x_p$ when $(n+1)p$ is near either 1 or $n$. An alternative approach is to fit a reasonable distribution function to the observations, as discussed in Section 3, and then estimate $x_p$ using Equation 7.32, where $F_X(x)$ is the fitted distribution.

Another simple and common approach to describing a distribution's centre, spread and shape is by reporting the moments of a distribution. The first moment about the origin is the *mean* of $X$ and is given by

$$\mu_X = \mathrm{E}[X] = \int_{-\infty}^{+\infty} x f_X(x) \mathrm{d}x \tag{7.34}$$

Moments other than the first are normally measured about the mean. The second moment measured about the mean is the *variance*, denoted Var($X$) or $\sigma_X^2$, where:

$$\sigma_X^2 = \mathrm{Var}(X) = \mathrm{E}[(X - \mu_X)^2] \tag{7.35}$$

The *standard deviation* $\sigma_X$ is the square root of the variance. While the mean $\mu_X$ is a measure of the central value of $X$, the standard deviation $\sigma_X$ is a measure of the spread of the distribution of $X$ about $\mu_X$.

Another measure of the variability in $X$ is the *coefficient of variation*,

$$CV_X = \frac{\sigma_X}{\mu_X} \tag{7.36}$$

The coefficient of variation expresses the standard deviation as a proportion of the mean. It is useful for comparing the relative variability of the flow in rivers of different sizes, or of rainfall variability in different regions when the random variable is strictly positive.

The third moment about the mean, denoted $\lambda_X$, measures the asymmetry, or *skewness*, of the distribution:

$$\lambda_X = \mathrm{E}[(X - \mu_X)^3] \tag{7.37}$$

Typically, the dimensionless coefficient of skewness $\gamma_X$ is reported rather than the third moment $\lambda_X$. The coefficient of skewness is the third moment rescaled by the cube of the standard deviation so as to be dimensionless and hence unaffected by the scale of the random variable:

$$\gamma_X = \frac{\lambda_X}{\sigma_X^3} \tag{7.38}$$

Streamflows and other natural phenomena that are necessarily non-negative often have distributions with positive skew coefficients, reflecting the asymmetric shape of their distributions.

When the distribution of a random variable is not known, but a set of observations $\{x_1, \ldots, x_n\}$ is available, the moments of the unknown distribution of $X$ can be estimated based on the sample values using the following equations. The sample estimate of the mean:

$$\bar{X} = \sum_{i=1}^{n} X_i/n \tag{7.39a}$$

The sample estimate of the variance:

$$\hat{\sigma}_X^2 = S_X^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2 \tag{7.39b}$$

The sample estimate of skewness:

$$\hat{\lambda}_X = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (X_i - \bar{X})^3 \tag{7.39c}$$

The sample estimate of the coefficient of variation:

$$\hat{C}V_X = S_X/\bar{X} \tag{7.39d}$$

The sample estimate of the coefficient of skewness:

$$\hat{\gamma}_X = \hat{\lambda}_X / S_X^3 \tag{7.39e}$$

The sample estimate of the mean and variance are often denoted as $\bar{x}$ and $s_x^2$ where the lower case letters are used when referring to a specific sample. All of these sample estimators provide only estimates of actual or true values. Unless the sample size $n$ is very large, the difference between the estimators and the true values of $\mu_X$, $\sigma_X^2$, $\lambda_X$, $CV_X$, and $\gamma_X$ may be large. In many ways, the field of statistics is about the precision of estimators of different quantities. One wants to know how well the mean of twenty annual rainfall depths describes the true expected annual rainfall depth, or how large the difference between the estimated 100-year flood and the true 100-year flood is likely to be.

As an example of the calculation of moments, consider the flood data in Table 7.2. These data have the following sample moments:

$$\bar{x} = 1549.2$$
$$s_X = 813.5$$
$$CV_X = 0.525$$
$$\hat{\gamma}_X = 0.712$$

As one can see, the data are positively skewed and have a relatively large coefficient of variance.

When discussing the accuracy of sample estimates, two quantities are often considered, *bias* and *variance*. An estimator $\hat{\theta}$ of a known or unknown quantity $\theta$ is a function of the observed values of the random variable $X$, say in $n$ different time periods, $X_1, \ldots, X_n$, that will be available to estimate the value of $\theta$; $\hat{\theta}$ may be written $\hat{\theta}[X_1, X_2, \ldots, X_n]$ to emphasize that $\hat{\theta}$ itself is a random variable. Its value depends on the sample values of the random variable that will be observed. An estimator $\hat{\theta}$ of a quantity $\theta$ is biased if $E[\hat{\theta}] \neq \theta$ and unbiased if $E[\hat{\theta}] = \theta$. The quantity $\{E[\hat{\theta}] - \theta\}$ is generally called the *bias of the estimator*.

An unbiased estimator has the property that its expected value equals the value of the quantity to be estimated. The sample mean is an unbiased estimate of the population mean $\mu_X$ because

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu_X \tag{7.40}$$

The estimator $S_X^2$ of the variance of $X$ is an unbiased estimator of the true variance $\sigma_X^2$ for independent observations (Benjamin and Cornell, 1970):

$$E\left[S_X^2\right] = \sigma_X^2 \tag{7.41}$$

However, the corresponding estimator of the standard deviation, $S_X$, is in general a biased estimator of $\sigma_X$ because

| date | discharge $m^3/s$ | date | discharge $m^3/s$ |
|---|---|---|---|
| 1930 | 410 | 1951 | 3070 |
| 1931 | 1150 | 1952 | 2360 |
| 1932 | 899 | 1953 | 1050 |
| 1933 | 420 | 1954 | 1900 |
| 1934 | 3100 | 1955 | 1130 |
| 1935 | 2530 | 1956 | 674 |
| 1936 | 758 | 1957 | 683 |
| 1937 | 1220 | 1958 | 1500 |
| 1938 | 1330 | 1959 | 2600 |
| 1939 | 1410 | 1960 | 3480 |
| 1940 | 3100 | 1961 | 1430 |
| 1941 | 2470 | 1962 | 809 |
| 1942 | 929 | 1963 | 1010 |
| 1943 | 586 | 1964 | 1510 |
| 1944 | 450 | 1965 | 1650 |
| 1946 | 1040 | 1966 | 1880 |
| 1947 | 1470 | 1967 | 1470 |
| 1948 | 1070 | 1968 | 1920 |
| 1949 | 2050 | 1969 | 2530 |
| 1950 | 1430 | 1970 | 1490 |

* Value for 1945 is missing.

E021101b

**Table 7.2.** Annual maximum discharges on Magra River, Italy, at Calamazza, 1930 – 70*.

$$E[S_X] \neq \sigma_X \tag{7.42}$$

The second important statistic often used to assess the accuracy of an estimator $\hat{\theta}$ is the variance of the estimator Var $\hat{\theta}$, which equals $E\{(\hat{\theta} - E[\hat{\theta}])^2\}$. For the mean of a set of independent observations, the variance of the sample mean is:

$$\text{Var}(\bar{X}) = \frac{\sigma_X^2}{n} \tag{7.43}$$

It is common to call $\sigma_x/\sqrt{n}$ the *standard error* of $\hat{x}$ rather than its standard deviation. The standard error of an average is the most commonly reported measure of its precision.

The bias measures the difference between the average value of an estimator and the quantity to be estimated.

The variance measures the spread or width of the estimator's distribution. Both contribute to the amount by which an estimator deviates from the quantity to be estimated. These two errors are often combined into the *mean square error*. Understanding that $\theta$ is fixed and the estimator $\hat{\theta}$ is a random variable, the mean squared error is the expected value of the squared distance (error) between $\theta$ and its estimator $\hat{\theta}$:

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= \text{E}[(\hat{\theta} - \theta)^2] = \text{E}\{[\hat{\theta} - \text{E}(\hat{\theta})] + [\text{E}(\hat{\theta}) - \theta]\}^2 \\ &= [\text{Bias}]^2 + \text{Var}(\hat{\theta}) \end{aligned} \tag{7.44}$$

where [Bias] is $\text{E}(\hat{\theta}) - \theta$.

Equation 7.44 shows that the MSE, equal to the expected average squared deviation of the estimator $\hat{\theta}$ from the true value of the parameter $\theta$, can be computed as the bias squared plus the variance of the estimator. MSE is a convenient measure of how closely $\hat{\theta}$ approximates $\theta$ because it combines both bias and variance in a logical way.

Estimation of the coefficient of skewness $\gamma_X$ provides a good example of the use of the MSE for evaluating the total deviation of an estimate from the true population value. The sample estimate $\hat{\gamma}_X$ of $\gamma_X$ is often biased, has a large variance, and its absolute value was shown by Kirby (1974) to be bounded by the square root of the sample size $n$:

$$|\hat{\gamma}_X| \leq \sqrt{n} \tag{7.45}$$

The bounds do not depend on the true skew, $\gamma_X$. However, the bias and variance of $\hat{\gamma}_X$ do depend on the sample size and the actual distribution of $X$. Table 7.3 contains the expected value and standard deviation of the estimated coefficient of skewness $\hat{\gamma}_X$ when $X$ has either a normal distribution, for which $\gamma_X = 0$, or a gamma distribution with $\gamma_X = 0.25$, 0.50, 1.00, 2.00, or 3.00. These values are adapted from Wallis et al. (1974 a,b) who employed moment estimators slightly different than those in Equation 7.39.

For the normal distribution, $\text{E}[\hat{\gamma}] = 0$ and $\text{Var}[\hat{\gamma}_X] \cong 5/n$. In this case, the skewness estimator is unbiased but highly variable. In all the other cases in Table 7.3, the skewness estimator is biased.

To illustrate the magnitude of these errors, consider the mean square error of the skew estimator $\hat{\gamma}_X$ calculated from a sample of size 50 when $X$ has a gamma distribution with $\gamma_X = 0.50$, a reasonable value for annual streamflows. The expected value of $\hat{\gamma}_X$ is 0.45; its variance equals $(0.37)^2$, its standard deviation is squared. Using Equation 7.44, the mean square error of $\hat{\gamma}_X$ is:

$$\begin{aligned} \text{MSE}(\hat{\gamma}_X) &= (0.45 - 0.50)^2 + (0.37)^2 \\ &= 0.0025 + 0.1369 = 0.139 \cong 0.14 \end{aligned} \tag{7.46}$$

An unbiased estimate of $\gamma_X$ is simply $(0.50/0.45)\,\hat{\gamma}_X$. Here the estimator provided by Equation 7.39e has been scaled to eliminate bias. This unbiased estimator has a mean squared error of:

$$\begin{aligned} \text{MSE}\left(\frac{0.50\hat{\gamma}_X}{0.48}\right) &= (0.50 - 0.50)^2 + \left[\left(\frac{0.50}{0.45}\right)(0.37)\right]^2 \\ &= 0.169 \cong 0.17 \end{aligned} \tag{7.47}$$

The mean square error of the unbiased estimator of $\hat{\gamma}_X$ is larger than the mean square error of the biased estimate. Unbiasing $\hat{\gamma}_X$ results in a larger mean square error for all the cases listed in Table 7.3 except for the normal distribution for which $\gamma_X = 0$, and the gamma distribution with $\gamma_X = 3.00$.

As shown here for the skew coefficient, biased estimators often have smaller mean square errors than unbiased estimators. Because the mean square error measures the total average deviation of an estimator from the quantity being estimated, this result demonstrates that the strict or unquestioning use of unbiased estimators is not advisable. Additional information on the sampling distribution of quantiles and moments is contained in Stedinger et al. (1993).

## 2.4. L-Moments and Their Estimators

L-moments are another way to summarize the statistical properties of hydrological data based on linear combinations of the original observations (Hosking, 1990). Recently, hydrologists have found that regionalization methods (to be discussed in Section 5) using L-moments are superior to methods using traditional moments (Hosking and Wallis, 1997; Stedinger and Lu, 1995). L-moments have also proved useful for construction of goodness-of-fit tests (Hosking et al., 1985; Chowdhury et al., 1991; Fill and Stedinger, 1995), measures of regional homogeneity and distribution selection methods (Vogel and Fennessey, 1993; Hosking and Wallis, 1997).

E021101c

| expected value of $\hat{\gamma}_X$ | | | | |
|---|---|---|---|---|
| distribution of $X$ | sample size | | | |
| | 10 | 20 | 50 | 80 |
| normal $\gamma_X = 0$ | 0.00 | 0.00 | 0.00 | 0.00 |
| gamma $\gamma_X = 0.25$ | 0.15 | 0.19 | 0.23 | 0.23 |
| $\gamma_X = 0.50$ | 0.31 | 0.39 | 0.45 | 0.47 |
| $\gamma_X = 1.00$ | 0.60 | 0.76 | 0.88 | 0.93 |
| $\gamma_X = 2.00$ | 1.15 | 1.43 | 1.68 | 1.77 |
| $\gamma_X = 3.00$ | 1.59 | 1.97 | 2.32 | 2.54 |
| upper bound on skew | 3.16 | 4.47 | 7.07 | 8.94 |

| standard deviation of $\hat{\gamma}_X$ | | | | |
|---|---|---|---|---|
| distribution of $X$ | sample size | | | |
| | 10 | 20 | 50 | 80 |
| normal $\gamma_X = 0$ | 0.69 | 0.51 | 0.34 | 0.26 |
| gamma $\gamma_X = 0.25$ | 0.69 | 0.52 | 0.35 | 0.28 |
| $\gamma_X = 0.50$ | 0.69 | 0.53 | 0.37 | 0.31 |
| $\gamma_X = 1.00$ | 0.70 | 0.57 | 0.44 | 0.38 |
| $\gamma_X = 2.00$ | 0.72 | 0.68 | 0.62 | 0.57 |
| $\gamma_X = 3.00$ | 0.74 | 0.76 | 0.77 | 0.77 |

**Table 7.3.** Sampling properties of coefficient of skewness estimator.

Source: Wallis et al. (1974b) who only divided by $n$ in the estimators of the moments, whereas in Equations 7.39b and 7.39c, we use the generally-adopted coefficients of $1/(n-1)$ and $n/(n-1)(n-2)$ for the variance and skew.

The first L-moment designated as $\lambda_1$ is simply the arithmetic mean:

$$\lambda_1 = \mathrm{E}[X] \tag{7.48}$$

Now let $X_{(i|n)}$ be the $i^{\text{th}}$ largest observation in a sample of size $n$ ($i = n$ corresponds to the largest). Then, for any distribution, the second L-moment, $\lambda_2$, is a description of scale based upon the expected difference between two randomly selected observations:

$$\lambda_2 = (1/2)\ \mathrm{E}[X_{(2|1)} - X_{(1|2)}] \tag{7.49}$$

Similarly, L-moment measures of skewness and kurtosis use three and four randomly selected observations, respectively.

$$\lambda_3 = (1/3)\ \mathrm{E}[X_{(3|3)} - 2X_{(2|3)} + X_{(1|3)}] \tag{7.50}$$

$$\lambda_4 = (1/4)\ \mathrm{E}[X_{(4|4)} - 3X_{(3|4)} + 3X_{(2|4)} - X_{(1|4)}] \tag{7.51}$$

Sample L-moment estimates are often computed using intermediate statistics called *probability weighted moments* (PWMs). The $r^{\text{th}}$ probability weighted moment is defined as:

$$\beta_r = \mathrm{E}\{X[F(X)]^r\} \tag{7.52}$$

where $F(X)$ is the cumulative distribution function of $X$. Recommended (Landwehr et al., 1979; Hosking and Wallis, 1995) unbiased PWM estimators, $b_r$, of $\beta_r$ are computed as:

$$b_0 = \bar{X}$$

$$b_1 = \frac{1}{n(n-1)} \sum_{j=2}^{n} (j-1) X_{(j)}$$

$$b_2 = \frac{1}{n(n-1)(n-2)} \sum_{j=3}^{n} (j-1)(j-2) X_{(j)} \tag{7.53}$$

These are examples of the general formula for computing estimators $b_r$ of $\widetilde{\beta}_r$.

$$b_r = \frac{1}{n}\sum_{j=r+1}^{n}\binom{j-1}{r}X_{(j)}\bigg/\binom{n-1}{r}$$
$$= \frac{1}{r+1}\sum_{j=r+1}^{n}\binom{j-1}{r}X_{(j)}\bigg/\binom{n}{r+1} \qquad (7.54)$$

for $r = 1, \ldots, n-1$.

L-moments are easily calculated in terms of PWMs using:

$$\lambda_1 = \beta_0$$
$$\lambda_2 = 2\beta_1 - \beta_0$$
$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$$
$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \qquad (7.55)$$

Wang (1997) provides formulas for directly calculating L-moment estimators of $\lambda_r$. Measures of the coefficient of variation, skewness and kurtosis of a distribution can be computed with L-moments, as they can with traditional product moments. Where skew primarily measures the asymmetry of a distribution, the kurtosis is an additional measure of the thickness of the extreme tails. Kurtosis is particularly useful for comparing symmetric distributions that have a skewness coefficient of zero. Table 7.4 provides definitions of the traditional coefficient of variation, coefficient of skewness and coefficient of kurtosis, as well as the L-moment, L-coefficient of variation, L-coefficient of skewness and L-coefficient of kurtosis.

The flood data in Table 7.2 can be used to provide an example of L-moments. Equation 7.53 yields estimates of the first three Probability Weighted Moments:

$$b_0 = 1{,}549.20$$
$$b_1 = 1003.89$$
$$b_2 = 759.02 \qquad (7.56)$$

Recall that $b_0$ is just the sample average $\bar{x}$. The sample L-moments are easily calculated using the probability weighted moments. One obtains:

$$\hat{\lambda}_1 = b_0 = 1{,}549$$
$$\hat{\lambda}_2 = 2b_1 - b_0 = 458$$
$$\hat{\lambda}_3 = 6b_2 - 6b_1 + b_0 = 80 \qquad (7.55)$$

Thus, the sample estimates of the L-coefficient of variation, $t_2$, and L-coefficient of skewness, $t_3$, are:

$$t_2 = 0.295$$
$$t_3 = 0.174 \qquad (7.58)$$

**Table 7.4. Definitions of dimensionless product-moment and L-moment ratios.**

E021101d

| name | common symbol | definition |
|---|---|---|
| **product-moment ratios** | | |
| coefficient of variation | $CV_X$ | $\sigma_X / \mu_X$ |
| skewness | $\gamma_X$ | $E[(X-\mu_X)^3]/\sigma_X^3$ |
| kurtosis | $\kappa_X$ | $E[(X-\mu_X)^4]/\sigma_X^4$ |
| **L-moment ratios** | | |
| L-coefficient of variation* | L-CV, $\tau_2$ | $\lambda_2 / \lambda_1$ |
| skewness | L-skewness, $\tau_3$ | $\lambda_3 / \lambda_2$ |
| kurtosis | L-kurtosis, $\tau_4$ | $\lambda_4 / \lambda_2$ |

* Hosking and Wallis (1997) use $\tau$ instead of $\tau_2$ to represent the L-CV ratio

# 3. Distributions of Random Events

A frequent task in water resources planning is the development of a model of some probabilistic or stochastic phenomena such as streamflows, flood flows, rainfall, temperatures, evaporation, sediment or nutrient loads, nitrate or organic compound concentrations, or water demands. This often requires one to fit a probability distribution function to a set of observed values of the random variable. Sometimes, one's immediate objective is to estimate a particular quantile of the distribution, such as the 100-year flood, 50-year six-hour-rainfall depth, or the minimum seven-day-average expected once-in-ten-year flow. Then the fitted distribution can supply an estimate of that quantity. In a stochastic simulation, fitted distributions are used to generate possible values of the random variable in question.

Rather than fitting a reasonable and smooth mathematical distribution, one could use the empirical distribution represented by the data to describe the possible values that a random variable may assume in the future and their frequency. In practice, the true mathematical form for the distribution that describes the events is not known. Moreover, even if it was, its functional form may have too many parameters to be of much practical use. Thus, using the empirical distribution represented by the data itself has substantial appeal.

Generally, the free parameters of the theoretical distribution are selected (estimated) so as to make the fitted distribution consistent with the available data. The goal is to select a physically reasonable and simple distribution to describe the frequency of the events of interest, to estimate that distribution's parameters, and ultimately to obtain quantiles, performance indices and risk estimates of satisfactory accuracy for the problem at hand. Use of a theoretical distribution has several advantages over use of the empirical distribution:

- It presents a smooth interpretation of the empirical distribution. As a result quantiles, performance indices and other statistics computed using the fitted distribution should be more accurate than those computed with the empirical distribution.
- It provides a compact and easy-to-use representation of the data.
- It is likely to provide a more realistic description of the range of values that the random variable may assume and their likelihood. For example, by using the empirical distribution, one implicitly assumes that no values larger or smaller than the sample maximum or minimum can occur. For many situations, this is unreasonable.
- Often one needs to estimate the likelihood of extreme events that lie outside the range of the sample (either in terms of $x$ values or in terms of frequency). Such extrapolation makes little sense with the empirical distribution.
- In many cases, one is not interested in the values of a random variable $X$, but instead in derived values of variables $Y$ that are functions of $X$. This could be a performance function for some system. If $Y$ is the performance function, interest might be primarily in its mean value E[$Y$], or the probability some standard is exceeded, Pr{$Y >$ standard}. For some theoretical $X$-distributions, the resulting $Y$-distribution may be available in closed form, thus making the analysis rather simple. (The normal distribution works with linear models, the lognormal distribution with product models, and the gamma distribution with queuing systems.)

This section provides a brief introduction to some useful techniques for estimating the parameters of probability distribution functions and for determining if a fitted distribution provides a reasonable or acceptable model of the data. Sub-sections are also included on families of distributions based on the normal, gamma and generalized-extreme-value distributions. These three families have found frequent use in water resources planning (Kottegoda and Rosso, 1997).

## 3.1. Parameter Estimation

Given a set of observations to which a distribution is to be fit, one first selects a distribution function to serve as a model of the distribution of the data. The choice of a distribution may be based on experience with data of that type, some understanding of the mechanisms giving rise to the data, and/or examination of the observations themselves. One can then estimate the parameters of the chosen distribution and determine if the fitted distribution provides an acceptable model of the data. A model is generally judged to be unacceptable if it is unlikely that one could have observed the available data were they actually drawn from the fitted distribution.

In many cases, good estimates of a distribution's parameters are obtained by the *maximum-likelihood-estimation* procedure. Give a set of $n$ *independent* observations $\{x_1, \dots, x_n\}$ of a continuous random variable $X$, the joint probability density function for the observations is:

$$f_{X_1,X_2,X_3,\dots,X_n}(x_1,\dots,x_n \mid \theta)$$
$$= f_X(x_1 \mid \theta) \cdot f_X(x_2 \mid \theta) \cdots f_X(x_n \mid \theta) \qquad (7.59)$$

where $\theta$ is the vector of the distribution's parameters.

The maximum likelihood estimator of $\theta$ is that vector $\theta$ which maximizes Equation 7.59 and thereby makes it as likely as possible to have observed the values $\{x_1, \dots, x_n\}$.

Considerable work has gone into studying the properties of maximum likelihood parameter estimates. Under rather general conditions, asymptotically the estimated parameters are normally distributed, unbiased and have the smallest possible variance of any asymptotically unbiased estimator (Bickel and Doksum, 1977). These, of course, are asymptotic properties, valid for large sample sizes $n$. Better estimation procedures, perhaps yielding biased parameter estimates, may exist for small sample sizes. Stedinger (1980) provides such an example. Still, maximum likelihood procedures are recommended with moderate and large samples, even though the iterative solution of nonlinear equations is often required.

An example of the maximum likelihood procedure for which closed-form expressions for the parameter estimates are obtained is provided by the lognormal distribution. The probability density function of a lognormally distributed random variable $X$ is:

$$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2\right\} \qquad (7.60)$$

Here, the parameters $\mu$ and $\sigma^2$ are the mean and variance of the logarithm of $X$, and not of $X$ itself.

Maximizing the logarithm of the joint density for $\{x_1, \dots, x_n\}$ is more convenient than maximizing the joint probability density itself. Hence, the problem can be expressed as the maximization of the *log-likelihood function*

$$\begin{aligned} L &= \ln \prod_{i=1}^{n} f[(x_i \mid \mu, \sigma)] \\ &= \sum_{i=1}^{n} \ln f(x_i \mid \mu, \sigma) \\ &= -\sum_{i=1}^{n} \ln(x_i\sqrt{2\pi}) - n\ln(\sigma) \\ &\quad - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[\ln(x_i)-\mu]^2 \end{aligned} \qquad (7.61)$$

The maximum can be obtained by equating to zero the partial derivatives $\partial L/\partial\mu$ and $\partial L/\partial\sigma$ whereby one obtains:

$$\begin{aligned} 0 &= \frac{\partial L}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}[\ln(x_i)-\mu] \\ 0 &= \frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}[\ln(x_i)-\mu]^2 \end{aligned} \qquad (7.62)$$

These equations yield the estimators

$$\begin{aligned} \hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n}\ln(x_i) \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}[\ln(x_i)-\hat{\mu}]^2 \end{aligned} \qquad (7.63)$$

The second-order conditions for a maximum are met and these values maximize Equation 7.59. It is useful to note that if one defines a new random variable $Y = \ln(X)$, then the maximum likelihood estimators of the parameters $\mu$ and $\sigma^2$, which are the mean and variance of the $Y$ distribution, are the sample estimators of the mean and variance of $Y$:

$$\begin{aligned} \hat{\mu} &= \bar{y} \\ \hat{\sigma}^2 &= [(n-1)/n]S_Y^2 \end{aligned} \qquad (7.64)$$

The correction $[(n-1)/n]$ in this last equation is often neglected.

The second commonly used parameter estimation procedure is the *method of moments*. The method of moments is often a quick and simple method for obtaining parameter estimates for many distributions. For a distribution with $m = 1$, 2 or 3 parameters, the first $m$ moments of the postulated distribution in Equations 7.34, 7.35 and 7.37 are equated to the estimates of those moments calculated using Equations 7.39. The resulting nonlinear equations are solved for the unknown parameters.

For the lognormal distribution, the mean and variance of $X$ as a function of the parameters $\mu$ and $\sigma$ are given by

$$\mu_X = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$
$$\sigma_X^2 = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1] \qquad (7.65)$$

Substituting $\bar{x}$ for $\mu_X$ and $s_X^2$ for $\sigma_X^2$ and solving for $\mu$ and $\sigma^2$ one obtains

$$\hat{\sigma}^2 = \ln\left(1 + s_X^2/\bar{x}^2\right)$$
$$\hat{\mu} = \ln\left(\frac{\bar{x}}{\sqrt{1 + s_X^2/\bar{x}^2}}\right) = \ln \bar{x} - \frac{1}{2}\hat{\sigma}^2 \qquad (7.66)$$

The data in Table 7.2 provide an illustration of both fitting methods. One can easily compute the sample mean and variance of the logarithms of the flows to obtain

$$\hat{\mu} = 7.202$$
$$\hat{\sigma}^2 = 0.3164 = (0.5625)^2 \qquad (7.67)$$

Alternatively, the sample mean and variance of the flows themselves are

$$\bar{x} = 1549.2$$
$$s_X^2 = 661{,}800 = (813.5)^2 \qquad (7.68)$$

Substituting those two values in Equation 7.66 yields

$$\hat{\mu} = 7.224$$
$$\sigma_X^2 = 0.2435 = (0.4935)^2 \qquad (7.69)$$

Method of moments and maximum likelihood are just two of many possible estimation methods. Just as method of moments equates sample estimators of moments to population values and solves for a distribution's parameters, one can simply equate L-moment estimators to population values and solve for the parameters of a distribution. The resulting method of L-moments has received considerable attention in the hydrological literature (Landwehr et al., 1978; Hosking et al., 1985; Hosking and Wallis, 1987; Hosking, 1990; Wang, 1997). It has been shown to have significant advantages when used as a basis for regionalization procedures that will be discussed in Section 5 (Lettenmaier et al., 1987; Stedinger and Lu, 1995; Hosking and Wallis, 1997).

Bayesian procedures provide another approach that is related to maximum likelihood estimation. Bayesian inference employs the likelihood function to represent the information in the data. That information is augmented with a prior distribution that describes what is known about constraints on the parameters and their likely values beyond the information provided by the recorded data available at a site. The likelihood function and the prior probability density function are combined to obtain the probability density function that describes the *posterior distribution* of the parameters:

$$f_\theta(\theta \mid x_1, x_2, \ldots, x_n) \propto f_X(x_1, x_2, \ldots, x_n \mid \theta)\xi(\theta) \qquad (7.70)$$

The symbol $\propto$ means 'proportional to' and $\xi(\theta)$ is the probability density function for the prior distribution for $\theta$ (Kottegoda and Rosso, 1997). Thus, except for a constant of proportionality, the probability density function describing the posterior distribution of the parameter vector $\theta$ is equal to the product of the likelihood function $f_X(x_1, x_2, \ldots, x_n \mid \theta)$ and the probability density function for the prior distribution $\xi(\theta)$ for $\theta$.

Advantages of the Bayesian approach are that it allows the explicit modelling of uncertainty in parameters (Stedinger, 1997; Kuczera, 1999) and provides a theoretically consistent framework for integrating systematic flow records with regional and other hydrological information (Vicens et al., 1975; Stedinger, 1983; Kuczera, 1983). Martins and Stedinger (2000) illustrate how a prior distribution can be used to enforce realistic constraints upon a parameter as well as providing a description of its likely values. In their case, use of a prior of the shape parameter $\kappa$ of a generalized extreme value (GEV) distribution (discussed in Section 3.6) allowed definition of generalized maximum likelihood estimators that, over the $\kappa$-range of interest, performed substantially better than maximum likelihood, moment, and L-moment estimators.

While Bayesian methods have been available for decades, the computational challenge posed by the solution of Equation 7.70 has been an obstacle to their use. Solutions to Equation 7.70 have been available for special cases such as normal data, and binomial and Poisson samples (Raiffa and Schlaifer, 1961; Benjamin and Cornell, 1970; Zellner, 1971). However, a new and very general set of Markov Chain Monte Carlo (MCMC) procedures (discussed in Section 7.2) allows numerical computation of the posterior distributions of parameters for a very broad class of models (Gilks et al., 1996). As a result, Bayesian methods are now becoming much more popular and are the standard approach for many difficult problems that are not easily addressed by traditional methods (Gelman et al., 1995; Carlin and Louis, 2000). The use of Monte Carlo Bayesian methods in flood frequency analysis, rainfall–runoff modelling, and evaluation of environmental pathogen concentrations are illustrated by Wang (2001), Bates and Campbell (2001) and Crainiceanu et al. (2002), respectively.

Finally, a simple method of fitting flood frequency curves is to plot the ordered flood values on special probability paper and then to draw a line through the data (Gumbel, 1958). Even today, that simple method is still attractive when some of the smallest values are zero or unusually small, or have been censored as will be discussed in Section 4 (Kroll and Stedinger, 1996). Plotting the ranked annual maximum series against a probability scale is always an excellent and recommended way to see what the data look like and for determining whether or not a fitted curve is consistent with the data (Stedinger et al., 1993).

Statisticians and hydrologists have investigated which of these methods most accurately estimates the parameters themselves or the quantiles of the distribution. One also needs to determine how accuracy should be measured. Some studies have used average squared deviations, some have used average absolute weighted deviations with different weights on under and over-estimation, and some have used the squared deviations of the log-quantile estimator (Slack et al., 1975; Kroll and Stedinger, 1996). In almost all cases, one is also interested in the bias of an estimator, which is the average value of the estimator minus the true value of the parameter or quantile being estimated. Special estimators have been developed to compute design events that on average are exceeded with the specified probability and have the anticipated risk of being exceeded (Beard, 1960, 1997; Rasmussen and Rosbjerg, 1989, 1991a,b; Stedinger, 1997; Rosbjerg and Madsen, 1998).

## 3.2. Model Adequacy

After estimating the parameters of a distribution, some check of model adequacy should be made. Such checks vary from simple comparisons of the observations with the fitted model (using graphs or tables) to rigorous statistical tests. Some of the early and simplest methods of parameter estimation were graphical techniques. Although quantitative techniques are generally more accurate and precise for parameter estimation, graphical presentations are invaluable for comparing the fitted distribution with the observations for the detection of systematic or unexplained deviations between the two. The observed data will plot as a straight line on probability graph paper if the postulated distribution is the true distribution of the observation. If probability graph paper does not exist for the particular distribution of interest, more general techniques can be used.

Let $x_{(i)}$ be the $i$th largest value in a set of observed values $\{x_i\}$ so that $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. The random variable $X_{(i)}$ provides a reasonable estimate of the $p$th quantile $x_p$ of the true distribution of $X$ for $p = i/(n+1)$. In fact, when one considers the cumulative probability $U_i$ associated with the random variable $X_{(i)}$, $U_i = F_X(X_{(i)})$, and if the observations $X_{(i)}$ are independent, then the $U_i$ have a beta distribution (Gumbel, 1958) with probability density function:

$$f_{U_i}(u) = \frac{n!}{(i-1)!\,(n-1)!}\, u^{i-1}(1-u)^{n-i} \qquad 0 \leq u \leq 1 \tag{7.71}$$

This beta distribution has mean

$$\mathrm{E}[U_i] = \frac{i}{n+1} \tag{7.72a}$$

and variance

$$\mathrm{Var}(U_i) = \frac{i(n-i+1)}{(n+1)^2\,(n+2)} \tag{7.72b}$$

A good graphical check of the adequacy of a fitted distribution $G(x)$ is obtained by plotting the observations $x_{(i)}$ versus $G^{-1}[i/(n+1)]$ (Wilk and Gnanadesikan, 1968). Even if $G(x)$ equalled to an exact degree the true $X$-distribution $F_X[x]$, the plotted points would not fall exactly on a 45° line through the origin of the graph. This would only occur if $F_X[x_{(i)}]$ exactly equalled $i/(n+1)$, and therefore each $x_{(i)}$ exactly equalled $F_X^{-1}[i/(n+1)]$.

An appreciation for how far an individual observation $x_{(i)}$ can be expected to deviate from $G^{-1}[i/(n+1)]$ can be obtained by plotting $G^{-1}[u_i^{(0.75)}]$ and $G^{-1}[u_i^{(0.25)}]$, where $u_i^{(0.75)}$ and $u_i^{(0.25)}$ are the upper and lower quartiles of the distribution of $U_i$ obtained from integrating the probability density function in Equation 7.71. The required incomplete beta function is also available in many software packages, including Microsoft Excel. Stedinger et al. (1993) show that $u_{(1)}$ and $(1 - u_{(n)})$ fall between $0.052/n$ and $3(n + 1)$ with a probability of 90%, thus illustrating the great uncertainty associated with the cumulative probability of the smallest value and the exceedance probability of the largest value in a sample.

Figures 7.2a and 7.2b illustrate the use of this *quantile–quantile plotting technique* by displaying the results of fitting a normal and a lognormal distribution to the annual maximum flows in Table 7.2 for the Magra River, Italy, at Calamazza for the years 1930 – 70. The observations of $X_{(i)}$, given in Table 7.2, are plotted on the vertical axis against the quantiles $G^{-1}[i/(n + 1)]$ on the horizontal axis.

A probability plot is essentially a scatter plot of the sorted observations $X_{(i)}$ versus some approximation of their expected or anticipated value, represented by $G^{-1}(p_i)$, where, as suggested, $p_i = i/(n + 1)$. The $p_i$ values are called *plotting positions*. A common alternative to $i/(n + 1)$ is $(i - 0.5)/n$, which results from a probabilistic interpretation of the empirical distribution of the data. Many reasonable plotting position formulas have been proposed based upon the sense in which $G^{-1}(p_i)$ should approximate $X_{(i)}$. The *Weibull formula* $i/(n + 1)$ and the *Hazen formula* $(i - 0.5)/n$ bracket most of the reasonable choices. Popular formulas are summarized by Stedinger et al. (1993), who also discuss the generation of probability plots for many distributions commonly employed in hydrology.

Rigorous statistical tests are available for trying to determine whether or not it is reasonable to assume that a given set of observations could have been drawn from a particular family of distributions. Although not the most powerful of such tests, the *Kolmogorov–Smirnov test* provides bounds within which every observation should lie if the sample is actually drawn from the assumed distribution. In particular, for $G = F_X$, the test specifies that

$$\Pr\left[G^{-1}\left(\frac{i}{n} - C_\alpha\right) \le X_{(i)} \le G^{-1}\left(\frac{i-1}{n} + C_\alpha\right)\forall i\right] = 1 - \alpha \tag{7.73}$$

where $C_\alpha$ is the critical value of the test at significance level $\alpha$. Formulas for $C_\alpha$ as a function of $n$ are contained in Table 7.5 for three cases: (1) when $G$ is completely

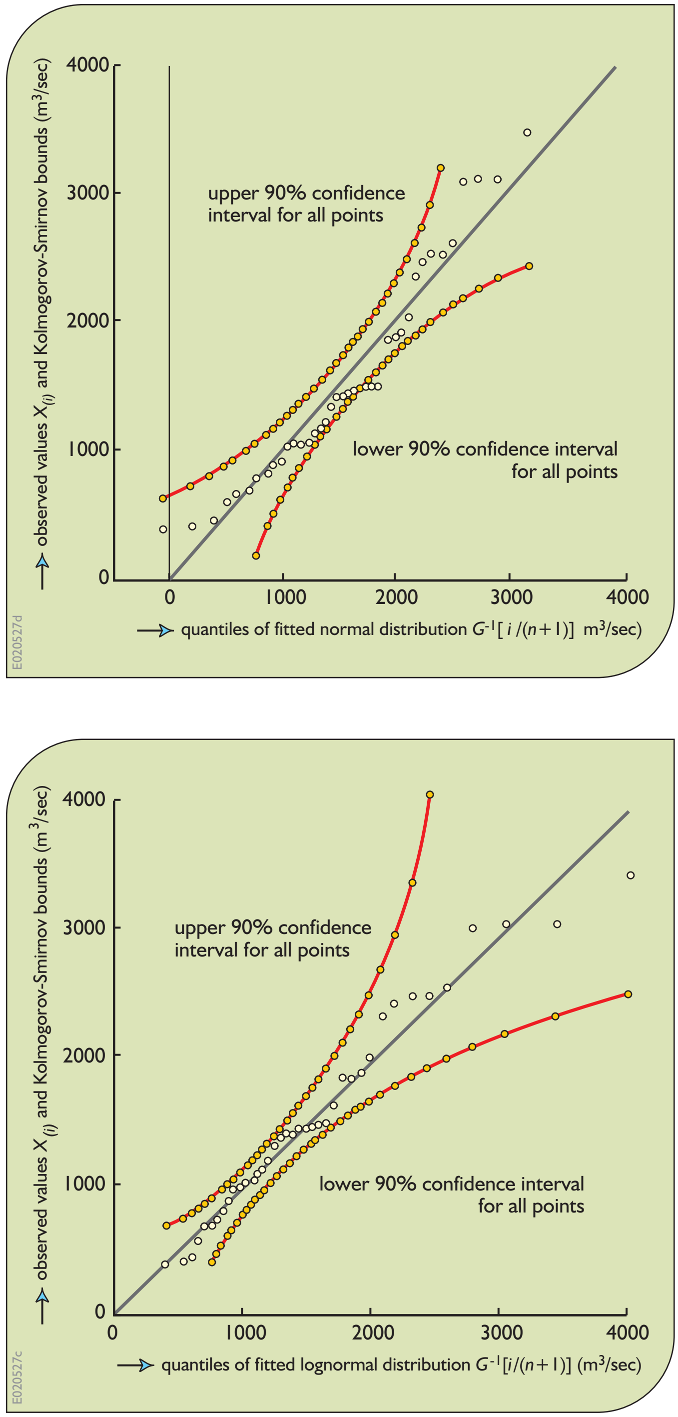



**Figure 7.2.** Plots of annual maximum discharges of Magra River, Italy, versus quantiles of fitted (a) normal and (b) lognormal distributions.

specified independent of the sample's values; (2) when $G$ is the normal distribution and the mean and variance are estimated from the sample with $\bar{x}$ and $s_X^2$; and (3) when $G$ is the exponential distribution and the scale parameter is estimated as $1/(\bar{x})$. Chowdhury et al. (1991) provide critical values for the Gumbel and generalized extreme value (GEV) distributions (Section 3.6) with known shape parameter $\kappa$. For other distributions, the values obtained from Table 7.5 may be used to construct approximate simultaneous confidence intervals for every $X_{(i)}$.

Figures 7.2a and b contain 90% confidence intervals for the plotted points constructed in this manner. For the normal distribution, the critical value of $C_\alpha$ equals $0.819/(\sqrt{n} - 0.01 + 0.85/\sqrt{n})$, where 0.819 corresponds to $\alpha = 0.10$. For $n = 40$, one computes $C_\alpha = 0.127$. As can be seen in Figure 7.2a, the annual maximum flows are not consistent with the hypothesis that they were drawn from a normal distribution; three of the observations lie outside the simultaneous 90% confidence intervals for all the points. This demonstrates a statistically significant lack of fit. The fitted normal distribution underestimates the quantiles corresponding to small and large probabilities while overestimating the quantiles in an intermediate range. In Figure 7.2b, deviations between the fitted lognormal distribution and the observations can be attributed to the differences between $F_X(x_{(i)})$ and $i/(n+1)$. Generally, the points are all near the 45° line through the origin, and no major systematic deviations are apparent.

The *Kolmogorov–Smirnov test* conveniently provides bounds within which every observation on a probability plot should lie if the sample is actually drawn from the assumed distribution, and thus is useful for visually evaluating the adequacy of a fitted distribution. However, it is not the most powerful test available for estimating which distribution a set of observations is likely to have been drawn from. For that purpose, several other more analytical tests are available (Filliben, 1975; Hosking, 1990; Chowdhury et al., 1991; Kottegoda and Rosso, 1997).

The *Probability Plot Correlation test* is a popular and powerful test of whether a sample has been drawn from a postulated distribution, though it is often weaker than alternative tests at rejecting thin-tailed alternatives (Filliben, 1975; Fill and Stedinger, 1995). A test with greater power has a greater probability of correctly determining that a sample is not from the postulated distribution. The Probability Plot Correlation Coefficient test employs the correlation $r$ between the ordered observations $x_{(i)}$ and the corresponding fitted quantiles $w_i = G^{-1}(p_i)$, determined by plotting positions $p_i$ for each $x_{(i)}$. Values of $r$ near 1.0 suggest that the observations could have been drawn from the fitted distribution: $r$ measures the linearity of the probability plot providing a quantitative assessment of fit. If $\bar{x}$ denotes the average value of the observations and $\bar{w}$ denotes the average value of the fitted quantiles, then

$$r = \frac{\sum (x_{(i)} - \bar{x})(w_i - \bar{w})}{\left[\left(\sum (x_{(i)} - \bar{x})^2 \sum (w_i - \bar{w})^2\right)\right]^{0.5}} \tag{7.74}$$

**Table 7.5. Critical values of Kolmogorov–Smirnov statistic as a function of sample size *n* (after Stephens, 1974).**

E021101e

| | significance level $\alpha$ | | | | |
|---|---|---|---|---|---|
| | 0.150 | 0.100 | 0.050 | 0.025 | 0.010 |
| $F_X$ completely specified: | | | | | |
| $C_\alpha(\sqrt{n} + 0.12 + 0.11/\sqrt{n})$ | 1.138 | 1.224 | 1.358 | 1.480 | 1.628 |
| $F_X$ normal with mean and variance estimated as $\bar{x}$ and $s_X^2$ | | | | | |
| $C_\alpha(\sqrt{n} + 0.01 + 0.85/\sqrt{n})$ | 0.775 | 0.819 | 0.895 | 0.995 | 1.035 |
| $F_X$ exponential with scale parameter b estimated as $1/(\bar{x})$ | | | | | |
| $(C_\alpha + 0.2/n)(\sqrt{n} + 0.26 + 0.5/\sqrt{n})$ | 0.926 | 0.990 | 1.094 | 1.190 | 1.308 |

values of $C_\alpha$ are calculated as follows:
for case 2 with $\alpha = 0.10$, $C_\alpha = 0.819/(\sqrt{n} - 0.01 + 0.85/\sqrt{n})$

Table 7.6 provides critical values for $r$ for the normal distribution, or the logarithms of lognormal variates, based upon the Blom plotting position that has $p_i = (i - 3/8)/(n + 1/4)$. Values for the Gumbel distribution are reproduced in Table 7.7 for use with the Gringorten plotting position $p_i = (i - 0.44)/(n + 0.12)$. The table also applies to logarithms of Weibull variates (Stedinger et al., 1993). Other tables are available for the GEV (Chowdhury et al., 1991), the Pearson type 3 (Vogel and McMartin, 1991), and exponential and other distributions (D'Agostino and Stephens, 1986).

Recently developed L-moment ratios appear to provide goodness-of-fit tests that are superior to both the Kolmogorov–Smirnov and the Probability Plot Correlation test (Hosking, 1990; Chowdhury et al., 1991; Fill and Stedinger, 1995). For normal data, the L-skewness estimator $\hat{\tau}_3$ (or $t_3$) would have mean zero and Var $\hat{\tau}_3 = (0.1866 + 0.8/n)/n$, allowing construction of a powerful test of normality against skewed alternatives using the normally distributed statistic

$$Z = t_3/\sqrt{(0.1866 + 0.8/n)/n} \qquad (7.75)$$

with a reject region $|Z| > z_{\alpha/2}$.

Chowdhury et al. (1991) derive the sampling variance of the L-CV and L-skewness estimators $\hat{\tau}_2$ and $\hat{\tau}_3$ as a function of $\kappa$ for the GEV distribution. These allow construction of a test of whether a particular data set is consistent with a GEV distribution with a regionally estimated value of $\kappa$, or a regional $\kappa$ and a regional coefficient of variation, CV. Fill and Stedinger (1995) show that the $\hat{\tau}_3$ L-skewness estimator provides a test for the Gumbel versus a general GEV distribution using the normally distributed statistic

$$Z = (\hat{\tau}_3 - 0.17)/\sqrt{(0.2326 + 0.70/n)/n} \qquad (7.76)$$

with a reject region $|Z| > z_{\alpha/2}$.

The literature is full of goodness-of-fit tests. Experience indicates that among the better tests there is often not a great deal of difference (D'Agostino and Stephens, 1986). Generation of a probability plot is most often a good idea because it allows the modeller to see what the data look like and where problems occur. The Kolmogorov–Smirnov test helps the eye

E021101f

| n | significance level 0.10 | 0.05 | 0.01 |
|---|---|---|---|
| 10 | 0.9347 | 0.9180 | 0.8804 |
| 15 | 0.9506 | 0.9383 | 0.9110 |
| 20 | 0.9600 | 0.9503 | 0.9290 |
| 30 | 0.9707 | 0.9639 | 0.9490 |
| 40 | 0.9767 | 0.9715 | 0.9597 |
| 50 | 0.9807 | 0.9764 | 0.9664 |
| 60 | 0.9835 | 0.9799 | 0.9710 |
| 75 | 0.9865 | 0.9835 | 0.9757 |
| 100 | 0.9893 | 0.9870 | 0.9812 |
| 300 | 0.99602 | 0.99525 | 0.99354 |
| 1,000 | 0.99854 | 0.99824 | 0.99755 |

**Table 7.6.** Lower critical values of the probability plot correlation test statistic for the normal distribution using $p_i = (i - 3/8)/(n + 1/4)$ (Vogel, 1987).

E021101g

| n | significance level 0.10 | 0.05 | 0.01 |
|---|---|---|---|
| 10 | 0.9260 | 0.9084 | 0.8630 |
| 20 | 0.9517 | 0.9390 | 0.9060 |
| 30 | 0.9622 | 0.9526 | 0.9191 |
| 40 | 0.9689 | 0.9594 | 0.9286 |
| 50 | 0.9729 | 0.9646 | 0.9389 |
| 60 | 0.9760 | 0.9685 | 0.9467 |
| 70 | 0.9787 | 0.9720 | 0.9506 |
| 80 | 0.9804 | 0.9747 | 0.9525 |
| 100 | 0.9831 | 0.9779 | 0.9596 |
| 300 | 0.9925 | 0.9902 | 0.9819 |
| 1,000 | 0.99708 | 0.99622 | 0.99334 |

**Table 7.7.** Lower critical values of the probability plot correlation test statistic for the Gumbel distribution using $p_i = (i - 0.44)/(n + 0.12)$ (Vogel, 1987).

interpret a probability plot by adding bounds to a graph, illustrating the magnitude of deviations from a straight line that are consistent with expected variability. One can also use quantiles of a beta distribution to illustrate the possible error in individual plotting positions, particularly at the extremes where that uncertainty is largest. The probability plot correlation test is a popular and powerful goodness-of-fit statistic. Goodness-of-fit tests based upon sample estimators of the L-skewness $\hat{\tau}_3$ for the normal and Gumbel distribution provide simple and useful tests that are not based on a probability plot.

## 3.3. Normal and Lognormal Distributions

The normal distribution and its logarithmic transformation, the lognormal distribution, are arguably the most widely used distributions in science and engineering. The probability density function of a normal random variable is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \quad \text{for} \quad -\infty < X < +\infty \tag{7.77}$$

where $\mu$ and $\sigma^2$ are equivalent to $\mu_X$ and $\sigma_X^2$, the mean and variance of $X$. Interestingly, the maximum likelihood estimators of $\mu$ and $\sigma^2$ are almost identical to the moment estimates $\bar{x}$ and $s_X^2$.

The normal distribution is symmetric about its mean $\mu_X$ and admits values from $-\infty$ to $+\infty$. Thus, it is not always satisfactory for modelling physical phenomena such as streamflows or pollutant concentrations, which are necessarily non-negative and have skewed distributions. A frequently used model for skewed distributions is the lognormal distribution. A random variable $X$ has a lognormal distribution if the natural logarithm of $X$, ln($X$), has a normal distribution. If $X$ is lognormally distributed, then by definition ln($X$) is normally distributed, so that the density function of $X$ is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2\right\} \frac{d(\ln x)}{dx} = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}[\ln(x/\eta)]^2\right\} \tag{7.78}$$

for $x > 0$ and $\mu = \ln(\eta)$. Here $\eta$ is the median of the $X$-distribution.

A lognormal random variable takes on values in the range $[0, +\infty]$. The parameter $\mu$ determines the scale of the $X$-distribution whereas $\sigma^2$ determines the shape of the distribution. The mean and variance of the lognormal distribution are given in Equation 7.65. Figure 7.3 illustrates the various shapes that the lognormal probability density function can assume. It is highly skewed with a thick right hand tail for $\sigma > 1$, and approaches a symmetric normal distribution as $\sigma \to 0$. The density function always has a value of zero at $x = 0$. The coefficient of variation and skew are:

$$CV_X = [\exp(\sigma^2) - 1]^{1/2} \tag{7.79}$$

$$\gamma_X = 3CV_X + CV_X^3 \tag{7.80}$$

The maximum likelihood estimates of $\mu$ and $\sigma^2$ are given in Equation 7.63 and the moment estimates in Equation 7.66. For reasonable-sized samples, the maximum likelihood estimates generally perform as well or better than the moment estimates (Stedinger, 1980).

The data in Table 7.2 were used to calculate the parameters of the lognormal distribution that would describe these flood flows. The results are reported in Equation 7.67. The two-parameter maximum likelihood and method of moments estimators identify parameter estimates for which the distribution skewness coefficients

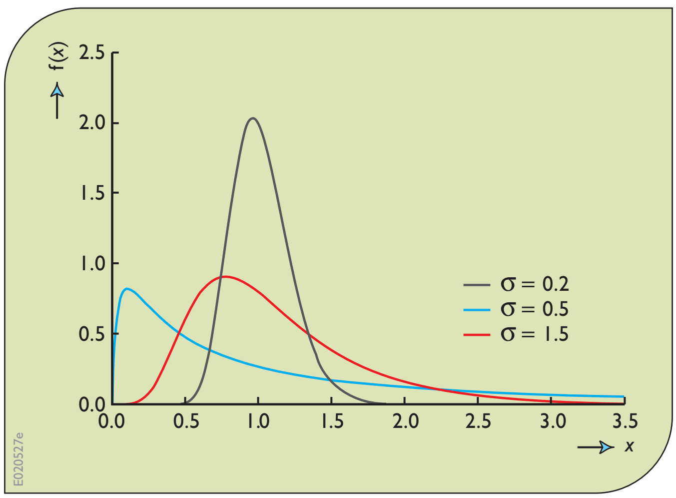



**Figure 7.3.** Lognormal probability density functions with various standard deviations $\sigma$.

are 2.06 and 1.72, which is substantially greater than the sample skew of 0.712.

A useful generalization of the two-parameter lognormal distribution is the shifted lognormal or three-parameter lognormal distribution obtained when $\ln(X - \tau)$ is described by a normal distribution, and $X \geq \tau$. Theoretically, $\tau$ should be positive if, for physical reasons, $X$ must be positive; practically, negative values of $\tau$ can be allowed when the resulting probability of negative values of $X$ is sufficiently small.

Unfortunately, maximum likelihood estimates of the parameters $\mu$, $\sigma^2$, and $\tau$ are poorly behaved because of irregularities in the likelihood function (Giesbrecht and Kempthorne, 1976). The method of moments does fairly well when the skew of the fitted distribution is reasonably small. A method that does almost as well as the moment method for low-skew distributions, and much better for highly skewed distributions, estimates $\tau$ by:

$$\hat{\tau} = \frac{x_{(1)}x_{(n)} - \hat{x}_{0.50}^2}{x_{(1)} + x_{(n)} - 2\hat{x}_{0.50}} \tag{7.81}$$

provided that $x_{(1)} + x_{(n)} - 2\hat{x}_{0.50} > 0$, where $x_{(1)}$ and $x_{(n)}$ are the smallest and largest observations and $\hat{x}_{0.50}$ is the sample median (Stedinger, 1980; Hoshi et al., 1984). If $x_{(1)} + x_{(n)} - 2\hat{x}_{0.50} < 0$, then the sample tends to be negatively skewed and a three-parameter lognormal distribution with a lower bound cannot be fit with this method. Good estimates of $\mu$ and $\sigma^2$ to go with $\hat{\tau}$ in Equation 7.81 are (Stedinger, 1980):

$$\hat{\mu} = \ln\left[\frac{\bar{x} - \hat{\tau}}{\sqrt{1 + s_X^2/(\bar{x} - \hat{\tau})^2}}\right]$$

$$\hat{\sigma}^2 = \ln\left[1 + \frac{s_X^2}{(\bar{x} - \hat{\tau})^2}\right] \tag{7.82}$$

For the data in Table 7.2, Equations 7.81 and 7.82 yield the hybrid moment-of-moments estimates of $\hat{\mu} = 7.606$, $\hat{\sigma}^2 = 0.1339 = (0.3659)^2$ and $\hat{\tau} = -600.1$ for the three-parameter lognormal distribution.

This distribution has a coefficient of skewness of 1.19, which is more consistent with the sample skewness estimator than were the values obtained when a two-parameter lognornal distribution was fit to the data. Alternatively, one can estimate $\mu$ and $\sigma^2$ by the sample mean and variance of $\ln(X - \hat{\tau})$ which yields the hybrid maximum likelihood estimates $\hat{\mu} = 7.605$, $\hat{\sigma}^2 = 0.1407 = (0.3751)^2$ and again $\hat{\tau} = -600.1$.

The two sets of estimates are surprisingly close in this instance. In this second case, the fitted distribution has a coefficient of skewness of 1.22.

Natural logarithms have been used here. One could have just as well use base 10 common logarithms to estimate the parameters; however, in that case the relationships between the log-space parameters and the real-space moments change slightly (Stedinger et al., 1993, Equation. 18.2.8).

## 3.4. Gamma Distributions

The gamma distribution has long been used to model many natural phenomena, including daily, monthly and annual streamflows as well as flood flows (Bobée and Ashkar, 1991). For a gamma random variable $X$,

$$f_X(x) = \frac{|\beta|}{\Gamma(\alpha)}(\beta x)^{\alpha-1}e^{-\beta x} \quad \beta x \geq 0$$

$$\mu_X = \frac{\alpha}{\beta}$$

$$\sigma_X^2 = \frac{\alpha}{\beta^2}$$

$$\gamma_X = \frac{2}{\sqrt{\alpha}} = 2CV_X \qquad \text{for } \beta > 0 \tag{7.83}$$

The gamma function, $\Gamma(\alpha)$, for integer $\alpha$ is $(\alpha - 1)!$. The parameter $\alpha > 0$ determines the shape of the distribution; $\beta$ is the scale parameter. Figure 7.4 illustrates the different shapes that the probability density function for a gamma variable can assume. As $\alpha \to \infty$, the gamma distribution approaches the symmetric normal distribution, whereas for $0 < \alpha < 1$, the distribution has a highly asymmetric J-shaped probability density function whose value goes to infinity as $x$ approaches zero.

The gamma distribution arises naturally in many problems in statistics and hydrology. It also has a very reasonable shape for such non-negative random variables as rainfall and streamflow. Unfortunately, its cumulative distribution function is not available in closed form, except for integer $\alpha$, though it is available in many software packages including Microsoft Excel. The gamma

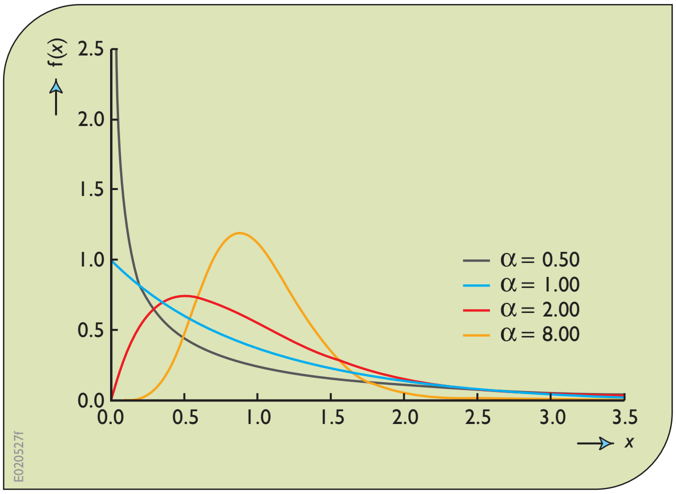



**Figure 7.4.** The gamma distribution function for various values of the shape parameter $\alpha$.

family includes a very special case: the exponential distribution is obtained when $\alpha = 1$.

The gamma distribution has several generalizations (Bobée and Ashkar, 1991). If a constant $\tau$ is subtracted from $X$ so that $(X - \tau)$ has a gamma distribution, the distribution of $X$ is a three-parameter gamma. This is also called a *Pearson type 3 distribution*, because the resulting distribution belongs to the third type of distributions suggested by the statistician Karl Pearson. Another variation is the log Pearson type 3 distribution obtained by fitting the logarithms of $X$ with a Pearson type 3 distribution. The log Pearson distribution is discussed further in the next section.

The method of moments may be used to estimate the parameters of the gamma distribution. For the three-parameter gamma distribution,

$$\hat{\tau} = \bar{x} - 2\left(\frac{s_X}{\hat{\gamma}_X}\right)$$

$$\hat{\sigma} = \frac{4}{(\hat{\gamma}_X)^2}$$

$$\hat{\beta} = \frac{2}{s_X \hat{\gamma}_X} \tag{7.84}$$

where $\bar{x}$, $s_X^2$, and $\gamma_X$ are estimates of the mean, variance, and coefficient of skewness of the distribution of $X$ (Bobée and Robitaille, 1977).

For the two-parameter gamma distribution,

$$\hat{\alpha} = \frac{(\bar{x})^2}{s_X^2}$$

$$\hat{\beta} = \frac{\bar{x}}{s_X^2} \tag{7.85}$$

Again, the flood record in Table 7.2 can be used to illustrate the different estimation procedures. Using the first three sample moments, one would obtain for the three-parameter gamma distribution the parameter estimates

$$\hat{\tau} = -735.6$$

$$\hat{\alpha} = 7.888$$

$$\hat{\beta} = 0.003452 = 1/427.2$$

Using only the sample mean and variance yields the method of moment estimators of the parameters of the two-parameter gamma distribution ($\tau = 0$),

$$\hat{\alpha} = 3.627$$

$$\hat{\beta} = 0.002341 = 1/427.2$$

The fitted two-parameter gamma distribution has a coefficient of skewness $\gamma$ of 1.05, whereas the fitted three-parameter gamma reproduces the sample skew of 0.712. As occurred with the three-parameter lognormal distribution, the estimated lower bound for the three-parameter gamma distribution is negative ($\hat{\tau} = -735.6$), resulting in a three-parameter model that has a smaller skew coefficient than was obtained with the corresponding two-parameter model. The reciprocal of $\hat{\beta}$ is also reported. While $\hat{\beta}$ has inverse $x$-units, $1/\hat{\beta}$ is a natural scale parameter that has the same units as $x$ and thus can be easier to interpret.

Studies by Thom (1958) and Matalas and Wallis (1973) have shown that maximum likelihood parameter estimates are superior to the moment estimates. For the two-parameter gamma distribution, Greenwood and Durand (1960) give approximate formulas for the maximum likelihood estimates (also Haan, 1977). However, the maximum likelihood estimators are often not used in practice because they are very sensitive to the smallest observations that sometimes suffer from measurement error and other distortions.

When plotting the observed and fitted quantiles of a gamma distribution, an approximation to the inverse of the distribution function is often useful. For $|\gamma| \leq 3$, the Wilson–Hilferty transformation

$$x_G = \mu + \sigma\left[\frac{2}{\gamma}\left(1 + \frac{\gamma x_N}{6} - \frac{\gamma^2}{36}\right)^3 - \frac{2}{\gamma}\right] \tag{7.86}$$

gives the quantiles $x_G$ of the gamma distribution in terms of $x_N$, the quantiles of the standard-normal distribution. Here $\mu$, $\sigma$, and $\gamma$ are the mean, standard deviation, and coefficient of skewness of $x_G$. Kirby (1972) and Chowdhury and Stedinger (1991) discuss this and other more complicated but more accurate approximations. Fortunately the availability of excellent approximations of the gamma cumulative distribution function and its inverse in Microsoft Excel and other packages has reduced the need for such simple approximations.

## 3.5. Log-Pearson Type 3 Distribution

The log-Pearson type 3 distribution (LP3) describes a random variable whose logarithms have a Pearson type 3 distribution. This distribution has found wide use in modelling flood frequencies and has been recommended for that purpose (IACWD, 1982). Bobée (1975) and Bobée and Ashkar (1991) discuss the unusual shapes that this hybrid distribution may take allowing negative values of $\beta$. The LP3 distribution has a probability density function given by

$$f_X(x) = \frac{|\beta|}{x\Gamma(\alpha)}[\beta(\ln(x) - \xi)]^{\alpha-1}\exp\{-\beta(\ln(x) - \xi)\} \tag{7.87}$$

with $\alpha > 0$, and $\beta$ either positive or negative. For $\beta < 0$, values are restricted to the range $0 < x < \exp(\xi)$. For $\beta > 0$, values have a lower bound so that $\exp(\xi) < X$. Figure 7.5 illustrates the probability density function for the LP3 distribution as a function of the skew $\gamma$ of the P3 distribution describing $\ln(X)$, with $\sigma_{\ln X} = 0.3$. The LP3 density function for $|\gamma| \leq 2$ can assume a wide range of shapes with both positive and negative skews. For $|\gamma| = 2$, the log-space P3 distribution is equivalent to an exponential distribution function, which decays exponentially as $x$ moves away from the lower bound ($\beta > 0$) or upper bound ($\beta < 0$): as a result the LP3 distribution has a similar shape. The space with $-1 < \gamma$ may be more realistic for describing variables whose probability density function becomes thinner as $x$ takes on large values. For $\gamma = 0$, the two-parameter lognormal distribution is obtained as a special case.

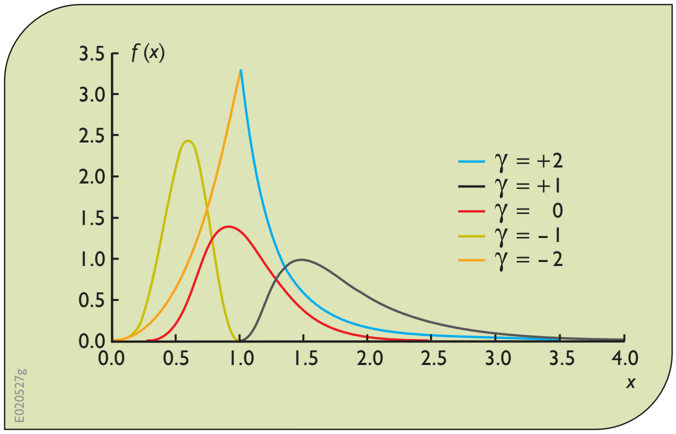



**Figure 7.5.** Log-Pearson type 3 probability density functions for different values of coefficient of skewness $\gamma$.

The LP3 distribution has mean and variance

$$\mu_X = e^{\xi}\left(\frac{\beta}{\beta-1}\right)^{\alpha}$$

$$\sigma_X^2 = e^{2\xi}\left\{\left(\frac{\beta}{\beta-2}\right)^{\alpha} - \left(\frac{\beta}{\beta-1}\right)^{2\alpha}\right\}$$

$$\text{for} \quad \beta > 2, \text{ or } \beta < 0. \tag{7.88}$$

For $0 < \beta < 2$, the variance is infinite.

These expressions are seldom used, but they do reveal the character of the distribution. Figures 7.6 and 7.7 provide plots of the real-space coefficient of skewness and coefficient of variation of a log-Pearson type 3 variate $X$ as a function of the standard deviation $\sigma_Y$ and coefficient of skew $\gamma_Y$ of the log-transformation $Y = \ln(X)$. Thus the standard deviation $\sigma_Y$ and skew $\gamma_Y$ of $Y$ are in log space. For $\gamma_Y = 0$, the log-Pearson type 3 distribution reduces to the two-parameter lognormal distribution discussed above, because in this case $Y$ has a normal distribution. For the lognormal distribution, the standard deviation $\sigma_Y$ serves as the sole shape parameter, and the coefficient of variation of $X$ for small $\sigma_Y$ is just $\sigma_Y$. Figure 7.7 shows that the situation is more

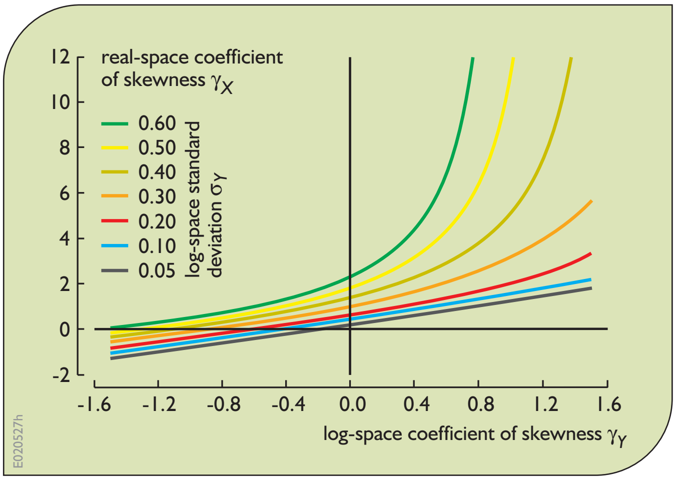



**Figure 7.6.** Real-space coefficient of skewness $\gamma_X$ for LP3 distributed $X$ as a function of log-space standard deviation $\sigma_Y$ and coefficient of skewness $\gamma_Y$ where $Y = \ln(X)$.

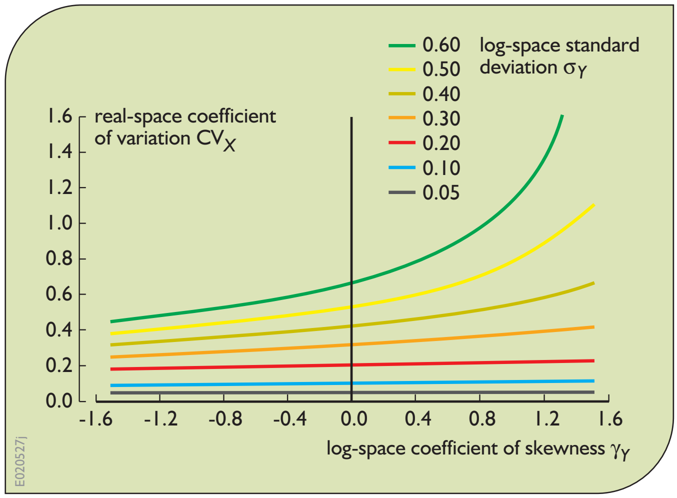



**Figure 7.7.** Real-space coefficient of variation $CV_X$ for LP3 distributed $X$ as a function of log-space standard deviation $\sigma_Y$ and coefficient of skewness $\gamma_Y$ where $Y = \ln(X)$.

complicated for the LP3 distribution. However, for small $\sigma_Y$, the coefficient of variation of $X$ is approximately $\sigma_Y$.

Again, the flood flow data in Table 7.2 can be used to illustrate parameter estimation. Using natural logarithms, one can estimate the log-space moments with the standard estimators in Equations 7.39 that yield:

$$\hat{\mu} = 7.202$$

$$\hat{\sigma} = 0.5625$$

$$\hat{\gamma} = -0.337$$

For the LP3 distribution, analysis generally focuses on the distribution of the logarithms $Y = \ln(X)$ of the flows, which would have a Pearson type 3 distribution with moments $\mu_Y$, $\sigma_Y$ and $\gamma_Y$ (IACWD, 1982; Bobée and Ashkar, 1991). As a result, flood quantiles are calculated as

$$x_p = \exp\{\mu_Y + \sigma_Y K_p[\gamma_Y]\} \tag{7.89}$$

where $K_p[\gamma_Y]$ is a frequency factor corresponding to cumulative probability $p$ for skewness coefficient $\gamma_Y$. ($K_p[\gamma_Y]$ corresponds to the quantiles of a three-parameter gamma distribution with zero mean, unit variance, and skewness coefficient $\gamma_Y$.)

Since 1967 the recommended procedure for flood frequency analysis by federal agencies in the United States has used this distribution. Current guidelines in Bulletin 17B (IACWD, 1982) suggest that the skew $\gamma_Y$ be estimated by a weighted average of the at-site sample skewness coefficient and a regional estimate of the skewness coefficient. Bulletin 17B also includes tables of frequency factors, a map of regional skewness estimators, checks for low outliers, confidence interval formula, a discussion of expected probability and a weighted-moments estimator for historical data.

## 3.6. Gumbel and GEV Distributions

The annual maximum flood is the largest flood flow during a year. One might expect that the distribution of annual maximum flood flows would belong to the set of extreme value distributions (Gumbel, 1958; Kottegoda and Rosso, 1997). These are the distributions obtained in the limit, as the sample size $n$ becomes large, by taking the largest of $n$ independent random variables. The Extreme Value (EV) type I distribution, or Gumbel distribution, has often been used to describe flood flows. It has the cumulative distribution function:

$$F_X(x) = \exp\{-\exp[-(x - \xi)/\alpha]\} \tag{7.90}$$

where $\xi$ is the location parameter. It has a mean and variance of

$$\mu_X = \xi + 0.5772\alpha$$

$$\sigma_X^2 = \pi^2\alpha^2/6 \cong 1.645\alpha^2 \tag{7.91}$$

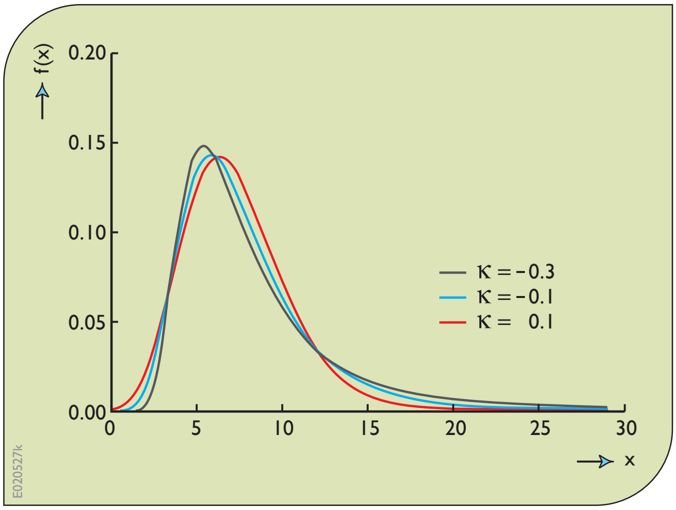



**Figure 7.8.** GEV density distributions for selected shape parameter $\kappa$ values.

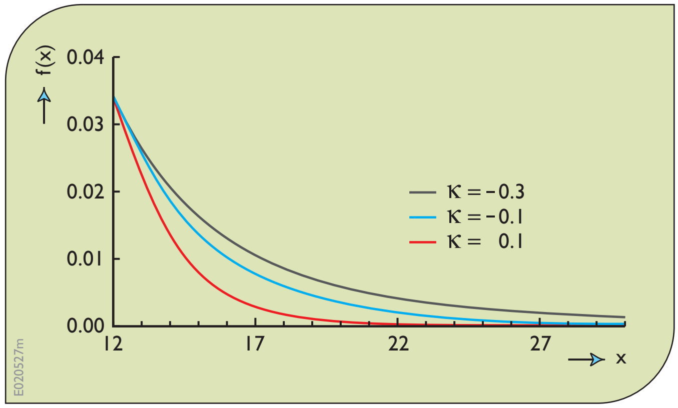



**Figure 7.9.** Right-hand tails of GEV distributions shown in Figure 7.8.

Its skewness coefficient has a fixed value equal to $\gamma_X = 1.1396$.

The generalized extreme value (GEV) distribution is a general mathematical expression that incorporates the type I, II, and III extreme value (EV) distributions for maxima (Gumbel, 1958; Hosking et al., 1985). In recent years it has been used as a general model of extreme events including flood flows, particularly in the context of regionalization procedures (NERC, 1975; Stedinger and Lu, 1995; Hosking and Wallis, 1997). The GEV distribution has the cumulative distribution function:

$$F_X(x) = \exp\{-[1 - \kappa(x-\xi)/\alpha]^{1/\kappa}\} \quad \text{for} \quad \kappa \neq 0 \qquad (7.92)$$

From Equation 7.92, it is clear that for $\kappa < 0$ (the typical case for floods), $x$ must exceed $\xi + \alpha/\kappa$, whereas for $\kappa > 0$, $x$ must be no greater than $\xi + \alpha/\kappa$ (Hosking and Wallis, 1987). The mean, variance, and skewness coefficient are (for $\kappa > -1/3$):

$$\mu_X = \xi + (\alpha/\kappa)\,[1 - \Gamma(1+\kappa)],$$

$$\sigma_X^2 = (\alpha/\kappa)^2\,\{\Gamma(1+2\kappa) - [\Gamma(1+\kappa)]^2\} \qquad (7.93)$$

$$\gamma_X = (\text{Sign } \kappa)\{-\Gamma(1+3\kappa) + 3\Gamma(1+\kappa)\,\Gamma(1+2\kappa) \\ -2[\Gamma(1+\kappa)]^3\}/\{\Gamma(1+2\kappa) - [\Gamma(1+\kappa)]^2\}^{3/2}$$

where $\Gamma(1+\kappa)$ is the classical gamma function. The Gumbel distribution is obtained when $\kappa = 0$. For $|\kappa| < 0.3$, the general shape of the GEV distribution is similar to the Gumbel distribution, though the right-hand tail is thicker for $\kappa < 0$, and thinner for $\kappa > 0$, as shown in Figures 7.8 and 7.9.

The parameters of the GEV distribution are easily computed using L-moments and the relationships (Hosking et al. (1985):

$$\kappa = 7.8590c + 2.9554c^2$$

$$\alpha = \kappa\lambda_2/[\Gamma(1+\kappa)(1 - 2^{-\kappa})] \qquad (7.94)$$

$$\xi = \lambda_1 + (\alpha/\kappa)[\Gamma(1+\kappa) - 1]$$

where

$$c = 2\lambda_2/(\lambda_3 + 3\lambda_2) - \ln(2)/\ln(3)$$
$$= [2/(\tau_3 + 3)] - \ln(2)/\ln(3)$$

As one can see, the estimator of the shape parameter $\kappa$ will depend only upon the L-skewness estimator $\hat{\tau}_3$. The estimator of the scale parameter $\alpha$ will then depend on the estimate of $\kappa$ and of $\lambda_2$. Finally, one must also use the sample mean $\lambda_1$ (Equation 7.48) to determine the estimate of the location parameter $\xi$.

Using the flood data in Table 7.2 and the sample L-moments computed in Section 2, one obtains first $c = -0.000896$ which yields $\hat{\kappa} = -0.007036$, $\hat{\xi} = 1{,}165.20$ and $\hat{\alpha} = 657.29$.

The small value of the fitted $\kappa$ parameter means that the fitted distribution is essentially a Gumbel distribution. Again, $\xi$ is a location parameter, not a lower bound, so its value resembles a reasonable $x$ value.

Madsen et al. (1997a) show that moment estimators can provide more precise quantile estimators. Martins and

Stedinger (2001b) found that with occasional uninformative samples, the MLE estimator of $\kappa$ could be entirely unrealistic resulting in absurd quantile estimators. However the use of a realistic prior distribution on $\kappa$ yielded generalized maximum likelihood estimators (GLME) that performed better than moment and L-moment estimators over the range of $\kappa$ of interest.

The generalized maximum likelihood estimators (GMLE) are obtained by maximizing the log-likelihood function, augmented by a prior density function on $\kappa$. A prior distribution that reflects general world-wide geophysical experience and physical realism is in the form of a beta distribution:

$$\pi(\kappa) = \Gamma(p)\,\Gamma(q)\,(0.5 + \kappa)^{p-1}\,(0.5 - \kappa)^{q-1}/\Gamma(p+q) \tag{7.95}$$

for $-0.5 < \kappa < +0.5$ with $p = 6$ and $q = 9$. Moreover, this prior assigns reasonable probabilities to the values of $\kappa$ within that range. For $\kappa$ outside the range $-0.4$ to $+0.2$ the resulting GEV distributions do not have density functions consistent with flood flows and rainfall (Martins and Stedinger, 2000). Other estimators implicitly have similar constraints. For example, L-moments restricts $\kappa$ to the range $\kappa > -1$, and the method of moments estimator employs the sample standard deviation so that $\kappa > -0.5$. Use of the sample skew introduces the constraint that $\kappa > -0.3$. Then given a set of independent observations $\{x_1, \dots, x_n\}$ drawn for a GEV distribution, the generalized likelihood function is:

$$\ln\{L(\xi,\alpha,\kappa \mid x_1,\dots,x_n)\} = -n\ln(\alpha) + \sum_{i=1}^{n}\left[\left(\frac{1}{\kappa}-1\right)\ln(y_i) - (y_i)^{1/\kappa}\right] + \ln[\pi(\kappa)]$$

with

$$y_i = [1-(\kappa/\alpha)(x_i-\xi)] \tag{7.96}$$

For feasible values of the parameters, $y_i$ is greater than 0 (Hosking et al., 1985). Numerical optimization of the generalized likelihood function is often aided by the additional constraint that $\min\{y_1, \dots, y_n\} \geq \varepsilon$ for some small $\varepsilon > 0$ so as to prohibit the search generating infeasible values of the parameters for which the likelihood function is undefined. The constraint should not be binding at the final solution.

The data in Table 7.2 again provide a convenient data set for illustrating parameter estimators. The L-moment estimators were used to generate an initial solution. Numerical optimization of the likelihood function in Equation 7.96 yielded the maximum likelihood estimators of the GEV parameters:

$$\hat{\kappa} = -0.0359, \quad \hat{\xi} = 1165.4 \text{ and } \hat{\alpha} = 620.2.$$

Similarly, use of the geophysical prior (Equation 7.95) yielded the generalized maximum likelihood estimators $\hat{\kappa} = -0.0823$, $\hat{\xi} = 1150.8$ and $\hat{\alpha} = 611.4$. Here the record length of forty years is too short to define reliably the shape parameter $\kappa$ so that result of using the prior is to pull $\kappa$ slightly toward the mean of the prior. The other two parameters adjust accordingly.

### 3.7. L-Moment Diagrams

Section 3 presented several families of distributions. The L-moment diagram in Figure 7.10 illustrates the relationships between the L-kurtosis ($\tau_4$) and L-skewness ($\tau_3$) for a number of distributions often used in hydrology. It shows that distributions with the same coefficient of skewness still differ in the thickness of their tails. This thickness is described by their kurtosis. Tail shapes are important if an analysis is sensitive to the likelihood of extreme events.

The normal and Gumbel distributions have a fixed shape and thus are presented by single points that fall on the Pearson type 3 (P3) curve for $\gamma = 0$, and the generalized extreme value (GEV) curve for $\kappa = 0$, respectively. The L-kurtosis/L-skewness relationships for the two-parameter and three-parameter gamma or P3 distributions are identical, as they are for the two-parameter and three-parameter lognormal distributions. This is because the addition of a location parameter does not change the range of fundamental shapes that can be generated. However, for the same skewness coefficient, the lognormal distribution has a larger kurtosis than the gamma or P3 distribution, and thus assigns larger probabilities to the largest events.

As the skewness of the lognormal and gamma distributions approaches zero, both distributions become normal and their kurtosis/skewness relationships merge. For the same L-skewness, the L-kurtosis of the GEV distribution is generally larger than that of the lognormal distribution. For positive $\kappa$ yielding almost symmetric or even negatively skewed GEV distributions, the GEV has a smaller kurtosis than the three-parameter lognormal distribution.

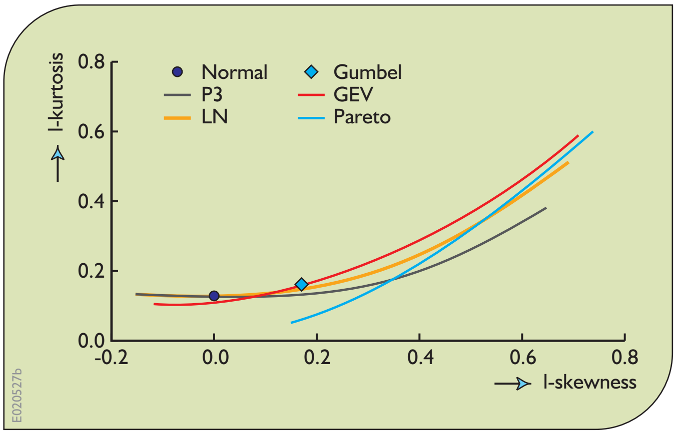



**Figure 7.10.** Relationships between L-skewness and L-kurtosis for various distributions.

The latter can be negatively skewed when the log normal location parameter $\tau$ is used as an upper bound.

Figure 7.10 also includes the three-parameter generalized Pareto distribution, whose cdf is:

$$F_X(x) = 1 - [1 - \kappa(x - \xi)/\alpha]^{1/\kappa} \tag{7.97}$$

(Hosking and Wallis, 1997). For $\kappa = 0$ it corresponds to the exponential distribution (gamma with $\alpha = 1$). This point is where the Pareto and P3 distribution L-kurtosis/L-skewness lines cross. The Pareto distribution becomes increasing more skewed for $\kappa < 0$, which is the range of interest in hydrology. The generalized Pareto distribution with $\kappa < 0$ is often used to describe peaks-over-a-threshold and other variables whose probability density function has its maximum at their lower bound. In that range for a given L-skewness, the Pareto distribution always has a larger kurtosis than the gamma distribution. In these cases the $\alpha$ parameter for the gamma distribution would need to be in the range $0 < \alpha < 1$, so that both distributions would be J-shaped.

As shown in Figure 7.10, the GEV distribution has a thicker right-hand tail than either the gamma/Pearson type 3 distribution or the lognormal distribution.

## 4. Analysis of Censored Data

There are many instances in water resources planning where one encounters censored data. A data set is censored if the values of observations that are outside a specified range of values are not specifically reported (David, 1981). For example, in water quality investigations many constituents have concentrations that are reported as $<T$, where $T$ is a reliable detection threshold (MacBerthouex and Brown, 2002). Thus the concentration of the water quality variable of interest was too small to be reliably measured. Likewise, low-flow observations and rainfall depths can be rounded to or reported as zero. Several approaches are available for analysis of censored data sets, including probability plots and probability-plot regression, conditional probability models and maximum likelihood estimators (Haas and Scheff, 1990; Helsel, 1990; Kroll and Stedinger, 1996; MacBerthouex and Brown, 2002).

Historical and physical paleoflood data provide another example of censored data. Before the beginning of a continuous measurement program on a stream or river, the stages of unusually large floods can be estimated on the basis of the memories of people who have experienced these events and/or physical markings in the watershed (Stedinger and Baker, 1987). Annual maximum floods that were not unusual were not recorded nor were they remembered. These missing data are censored data. They cover periods between occasional large floods that have been recorded or that have left some evidence of their occurrence (Stedinger and Cohn, 1986).

The discussion below addresses probability-plot methods for use with censored data. Probability-plot methods have a long history of use for this purpose because they are relatively simple to use and to understand. Moreover, recent research has shown that they are relatively efficient when the majority of values are observed, and unobserved values are known only to be below (or above) some detection limit or perception threshold that serves as a lower (or upper) bound. In such cases, probability-plot regression estimators of moments and quantiles are as accurate as maximum likelihood estimators. They are almost as good as estimators computed with complete samples (Helsel and Cohn, 1988; Kroll and Stedinger, 1996).

Perhaps the simplest method for dealing with censored data is adoption of a conditional probability model. Such models implicitly assume that the data are drawn from one of two classes of observations: those below a single threshold, and those above the threshold. This model is appropriate for simple cases where censoring occurs because small observations are recorded as 'zero,' as often happens with low-flow, low pollutant concentration, and some flood records. The conditional probability model introduces an extra parameter $P_0$ to describe the probability that an observation is 'zero'. If $r$ of a total of $n$ observations were observed because they exceeded the threshold, then $P_0$ is estimated as $(n - r)/n$. A continuous distribution $G_X(x)$ is derived for the strictly positive 'non-zero' values of $X$. Then the parameters of the $G$ distribution can be estimated using any procedure appropriate for complete uncensored samples. The unconditional cdf for any value $x > 0$, is then

$$F_X(x) = P_0 + (1 - P_0)\, G(x) \tag{7.98}$$

This model completely decouples the value of $P_0$ from the parameters that describe the $G$ distribution.

Section 7.2 discusses probability plots and plotting positions useful for graphical displays of data to allow a visual examination of the empirical frequency curve. Suppose that among $n$ samples a detection limit is exceeded by the observations $r$ times. The natural estimator of the exceedance probability $P_0$ of the perception threshold is again $(n - r)/n$. If the $r$ values that exceeded the threshold are indexed by $i = 1, \ldots, r$, wherein $x_{(r)}$ is the largest observation, reasonable plotting positions within the interval $[P_0, 1]$ are:

$$p_i = P_0 + (1 - P_0)\,[(i - a)/(r + 1 - 2a)] \tag{7.99}$$

where $a$ defines the plotting position that is used; $a = 0$ is reasonable (Hirsch and Stedinger, 1987). Helsel and Cohn (1988) show that reasonable choices for $a$ generally make little difference. Both papers discuss development of plotting positions when there are different thresholds, as occurs when the analytical precision of instrumentation changes over time. If there are many exceedances of the threshold so that $r >> (1 - 2a)$, $p_i$ is indistinguishable from

$$p_i' = [i + (n - r) - a]/(n + 1 - 2a) \tag{7.100}$$

where again, $i = 1, \ldots, r$. These values correspond to the plotting positions that would be assigned to the largest $r$ observations in a complete sample of $n$ values.

The idea behind the probability-plot regression estimators is to use the probability plot for the observed data to define the parameters of the whole distribution. And if a sample mean, sample variance or quantiles are needed, then the distribution defined by the probability plot is used to fill in the missing (censored) observations so that standard estimators of the mean, standard deviation and of quantiles can be employed. Such fill-in procedures are efficient and relatively robust for fitting a distribution and estimating various statistics with censored water quality data when a modest number of the smallest observations are censored (Helsel, 1990; Kroll and Stedinger, 1996).

Unlike the conditional probability approach, here the below threshold probability $P_0$ is linked with the selected probability distribution for the above-threshold observations. The observations below the threshold are censored but are in all other respects envisioned as coming from the same distribution that is used to describe the observed above-threshold values.

When water quality data are well described by a lognormal distribution, available values $\ln[X_{(1)}] \leq \cdots \leq \ln[X_{(r)}]$ can be regressed upon $F^{-1}[p_i] = \mu + \sigma F^{-1}[p_i]$ for $i = 1, \ldots, r$, where the $r$ largest observations in a sample of size $n$ are available. If regression yields constant $m$ and slope $s$ corresponding to the first and second population moments $\mu$ and $\sigma$, a good estimator of the $p$th quantile is

$$x_p = \exp[m + sz_p] \tag{7.101}$$

wherein $z_p$ is the $p^{\text{th}}$ quantile of the standard normal distribution. To estimate sample means and other statistics one can fill in the missing observations with

$$x(j) = \exp\{y(j)\} \quad \text{for} \quad j = 1, \ldots, (n - r) \tag{7.102}$$

where

$$y(j) = m + sF^{-1}\{P_0[(j - a)/(n - r + 1 - 2a)]\} \tag{7.103}$$

Once a complete sample is constructed, standard estimators of the sample mean and variance can be calculated, as can medians and ranges. By filling in the missing small observations, and then using complete-sample estimators of statistics of interest, the procedure is relatively insensitive to the assumption that the observations actually have a lognormal distribution.

Maximum likelihood estimators are quite flexible, and are more efficient than plotting-position methods when the values of the observations are not recorded because they are below or above the perception threshold (Kroll and Stedinger, 1996). Maximum likelihood methods allow the observations to be represented by exact values, ranges and various thresholds that either were or were not exceeded at various times. This can be particularly important with historical flood data sets because the magnitudes of many historical floods are not recorded precisely, and it may be known that a threshold was never crossed or was crossed at most once or twice in a long period (Stedinger and Cohn, 1986; Stedinger, 2000; O'Connell et al., 2002). Unfortunately, maximum likelihood estimators for the LP3 distribution have proven to be problematic. However, recently developed expected moment estimators seem to do as well as maximum likelihood estimators with the LP3 distribution (Cohn et al., 1997, 2001; Griffs et al., 2004).

While often a computational challenge, maximum likelihood estimators for complete samples, and samples with some observations censored, pose no conceptual challenge. One need only write the maximum likelihood function for the data and proceed to seek the parameter values that maximize that function. Thus if $F(x \mid \boldsymbol{\theta})$ and $f(x \mid \boldsymbol{\theta})$ are the cumulative distribution and probability density functions that should describe the data, and $\boldsymbol{\theta}$ is the vector of parameters of the distribution, then for the case described above wherein $x_1, \ldots, x_r$ are $r$ of $n$ observations that exceeded a threshold $T$, the likelihood function would be (Stedinger and Cohn, 1986):

$$\begin{aligned} &L(\boldsymbol{\theta} \mid r, n, x_1, \ldots, x_r) \\ &\quad = F(T \mid \boldsymbol{\theta})^{(n-r)} f(x_1 \mid \boldsymbol{\theta}) f(x_2 \mid \boldsymbol{\theta}) \cdots f(x_r \mid \boldsymbol{\theta}) \end{aligned} \tag{7.104}$$

Here, $(n - r)$ observations were below the threshold $T$, and the probability an observation is below $T$ is $F(T \mid \boldsymbol{\theta})$ which then appears in Equation 7.104 to represent that observation. In addition, the specific values of the $r$ observations $x_1, \ldots, x_r$ are available, where the probability an observation is in a small interval of width $\delta$ around $x_i$ is $\delta f(x_i \mid \boldsymbol{\theta})$. Thus strictly speaking the likelihood function also includes a term $\delta^r$. Here what is known of the magnitude of all of the $n$ observations is included in the likelihood function in the appropriate way. If all that were known of some observation was that it exceeded a threshold $M$, then that value should be represented by a term $[1 - F(M \mid \boldsymbol{\theta})]$ in the likelihood function. Similarly, if all that were known was that the value was between $L$ and $M$, then a term $[F(M \mid \boldsymbol{\theta}) - F(L \mid \boldsymbol{\theta})]$ should be included in the likelihood function. Different thresholds can be used to describe different observations corresponding to changes in the quality of measurement procedures. Numerical methods can be used to identify the parameter vector that maximizes the likelihood function for the data available.

## 5. Regionalization and Index-Flood Method

Research has demonstrated the potential advantages of 'index flood' procedures (Lettenmaier et al., 1987; Stedinger and Lu, 1995; Hosking and Wallis, 1997; Madsen, and Rosbjerg, 1997a). The idea behind the index-flood approach is to use the data from many hydrologically 'similar' basins to estimate a dimensionless flood distribution (Wallis, 1980). Thus this method 'substitutes space for time' by using regional information to compensate for having relatively short records at each site. The concept underlying the index-flood method is that the distributions of floods at different sites in a 'region' are the same except for a scale or index-flood parameter that reflects the size, rainfall and runoff characteristics of each watershed. Research is revealing when this assumption may be reasonable. Often a more sophisticated multi-scaling model is appropriate (Gupta and Dawdy, 1995a; Robinson and Sivapalan, 1997).

Generally the mean is employed as the index flood. The problem of estimating the $p^{\text{th}}$ quantile $x_p$ is then reduced to estimation of the mean, $\mu_x$, for a site and the ratio $x_p/\mu_x$ of the $p^{\text{th}}$ quantile to the mean. The mean can often be estimated adequately with the record available at a site, even if that record is short. The indicated ratio is estimated using regional information. The British Flood Studies Report (NERC, 1975) calls these normalized flood distributions "growth curves".

Key to the success of the index-flood approach is identification of sets of basins that have similar coefficients of variation and skew. Basins can be grouped geographically, as well as by physiographic characteristics including drainage area and elevation. Regions need not be geographically contiguous. Each site can potentially be assigned its own unique region consisting of sites with which it is particularly similar (Zrinji and Burn, 1994), or regional regression equations can be derived to compute normalized regional quantiles as a function of a site's physiographic characteristics and other statistics (Fill and Stedinger, 1998).

Clearly the next step for regionalization procedures, such as the index-flood method, is to move away from estimates of regional parameters that do not depend upon basin size and other physiographic parameters. Gupta et al. (1994) argue that the basic premise of the index-flood method – that the coefficient of variation of floods is relatively constant – is inconsistent with the known relationships between the coefficient of variation (CV) and drainage area (see also Robinson and Sivapalan, 1997). Recently, Fill and Stedinger (1998) built such a relationship into an index-flood procedure by using a regression model to explain variations in the normalized quantiles. Tasker and Stedinger (1986) illustrated how one might relate log-space skew to physiographic basin characteristics (see also Gupta and Dawdy, 1995b). Madsen and Rosbjerg (1997b) did the same for a regional model of $\kappa$ for the GEV distribution. In both studies, only a binary variable representing 'region' was found useful in explaining variations in these two shape parameters.

Once a regional model of alternative shape parameters is derived, there may be some advantage to combining such regional estimators with at-site estimators employing an empirical Bayesian framework or some other weighting schemes. For example, Bulletin 17B recommends weigh at-site and regional skewness estimators, but almost certainly places too much weight on the at-site values (Tasker and Stedinger, 1986). Examples of empirical Bayesian procedures are provided by Kuczera (1982), Madsen and Rosbjerg (1997b) and Fill and Stedinger (1998). Madsen and Rosbjerg's (1997b) computation of a $\kappa$-model with a New Zealand data set demonstrates how important it can be to do the regional analysis carefully, taking into account the cross-correlation among concurrent flood records.

When one has relatively few data at a site, the index-flood method is an effective strategy for deriving flood frequency estimates. However, as the length of the available record increases, it becomes increasingly advantageous to also use the at-site data to estimate the coefficient of variation as well. Stedinger and Lu (1995) found that the L-moment/GEV index-flood method did quite well for 'humid regions' ($CV \approx 0.5$) when $n < 25$, and for semi-arid regions ($CV \approx 1.0$) for $n < 60$, if reasonable care is taken in selecting the stations to be included in a regional analysis. However, with longer records, it became advantageous to use the at-site mean and L-CV with a regional estimator of the shape parameter for a GEV distribution. In many cases this would be roughly equivalent to fitting a Gumbel distribution corresponding to a shape parameter $\kappa = 0$. Gabriele and Arnell (1991) develop the idea of having regions of different sizes for different parameters. For realistic hydrological regions, these and other studies illustrate the value of regionalizing estimators of the shape, and often the coefficient of variation of a distribution.

## 6. Partial Duration Series

Two general approaches are available for modelling flood and precipitation series (Langbein, 1949). An annual maximum series considers only the largest event in each year. A partial duration series (PDS) or peaks-over-threshold (POT) approach includes all 'independent' peaks above a truncation or threshold level. An objection to using annual maximum series is that it employs only the largest event in each year, regardless of whether the second-largest event in a year exceeds the largest events of other years. Moreover, the largest annual flood flow in a dry year in some arid or semi-arid regions may be zero, or so small that calling them floods is misleading. When considering rainfall series or pollutant discharge events, one may be interested in modelling all events that occur within a year that exceed some threshold of interest.

Use of a partial duration series framework avoids such problems by considering all independent peaks that exceed a specified threshold. Furthermore, one can estimate annual exceedance probabilities from the analysis of partial duration series. Arguments in favour of partial duration series are that relatively long and reliable records are often available, and if the arrival rate for peaks over the threshold is large enough (1.65 events/year for the Poisson-arrival with exponential-exceedance model), partial duration series analyses should yield more accurate estimates of extreme quantiles than the corresponding annual-maximum frequency analyses (NERC, 1975; Rosbjerg, 1985). However, when fitting a three-parameter distribution, there seems to be little advantage from the use of a partial duration series approach over an annual maximum approach. This is true even when the partial duration series includes many more peaks than the maximum series because both contain the same largest events (Martins and Stedinger, 2001a).

A drawback of partial duration series analyses is that one must have criteria to identify only independent peaks (and not multiple peaks corresponding to the same event). Thus, such analysis can be more complicated than analyses using annual maxima. Partial duration models, perhaps with parameters that vary by season, are often used to estimate expected damages from hydrological events when more than one damage-causing event can occur in a season or within a year (North, 1980).

A model of a partial duration series has at least two components: first, one must model the arrival rate of events larger than the threshold level; second, one must model the magnitudes of those events. For example, a Poisson distribution has often been used to model the arrival of events, and an exponential distribution to describe the magnitudes of peaks that exceed the threshold.

There are several general relationships between the probability distribution for annual maximum and the frequency of events in a partial duration series. For a partial duration series model, let $\lambda$ be the average arrival rate of flood peaks greater than the threshold $x_0$ and let $G(x)$ be the probability that flood peaks, when they occur, are less than $x > x_0$, and thus those peaks fall in the range $[x_0, x]$. The annual exceedance probability for a flood, denoted $1/T_a$, corresponding to an annual return period $T_a$, is related to the corresponding exceedance probability $q_e = [1 - G(x)]$ for level $x$ in the partial duration series by

$$1/T_a = 1 - \exp\{-\lambda q_e\} = 1 - \exp\{-1/T_p\} \qquad (7.105)$$

where $T_p = 1/(\lambda q_e)$ is the average return period for level $x$ in the partial duration series.

Many different choices for $G(x)$ may be reasonable. In particular, the generalized Pareto distribution (GPD) is a simple distribution useful for describing floods that exceed a specified lower bound. The cumulative distribution function for the generalized three-parameter Pareto distribution is:

$$F_X(x) = 1 - [1 - \kappa(x - \xi)/\alpha]^{1/\kappa} \qquad (7.106)$$

with mean and variance

$$\mu_X = \xi + \alpha/(1+\kappa)\kappa$$

$$\sigma_X^2 = \alpha^2/[(1+\kappa)^2(1+2\kappa)] \qquad (7.107)$$

where for $\kappa < 0$, $\xi < x < \infty$, whereas for $\kappa > 0$, $\xi < x < \xi + \alpha/\kappa$ (Hosking and Wallis, 1987). A special case of the GPD is the two-parameter exponential distribution with $\kappa = 0$. Method of moment estimators work relatively well (Rosbjerg et al., 1992).

Use of a generalized Pareto distribution for $G(x)$ with a Poisson arrival model yields a GEV distribution for the annual maximum series greater than $x_0$ (Smith, 1984; Stedinger et al., 1993; Madsen et al., 1997a). The Poisson-Pareto and Poisson-GPD models provide very reasonable descriptions of flood risk (Rosbjerg et al., 1992). They have the advantage that they focus on the distribution of the larger flood events, and regional estimates of the GEV distribution's shape parameter $\kappa$ from annual maximum and partial duration series analyses can be used interchangeably.

Madsen and Rosbjerg (1997a) use a Poisson-GPD model as the basis of a partial duration series index-flood procedure. Madsen et al. (1997b) show that the estimators are fairly efficient. They pooled information from many sites to estimate the single shape parameter $\kappa$ and the arrival rate where the threshold was a specified percentile of the daily flow duration curve at each site. Then at-site information was used to estimate the mean above-threshold flood. Alternatively, one could use the at-site data to estimate the arrival rate as well.

## 7. Stochastic Processes and Time Series

Many important random variables in water resources are functions whose values change with time. Historical records of rainfall or streamflow at a particular site are a sequence of observations called a *time series*. In a time series, the observations are ordered by time, and it is generally the case that the observed value of the random variable at one time influences one's assessment of the distribution of the random variable at later times. This means that the observations are not independent. Time series are conceptualized as being a single observation of a *stochastic process,* which is a generalization of the concept of a random variable.

This section has three parts. The first presents the concept of *stationarity* and the basic statistics generally used to describe the properties of a stationary stochastic process. The second presents the definition of a Markov process and the Markov chain model. Markov chains are a convenient model for describing many phenomena, and are often used in synthetic flow and rainfall generation and optimization models. The third part discusses the sampling properties of statistics used to describe the characteristics of many time series.

## 7.1. Describing Stochastic Processes

A random variable whose value changes through time according to probabilistic laws is called a *stochastic process*. An observed *time series* is considered to be one realization of a stochastic process, just as a single observation of a random variable is one possible value the random variable may assume. In the development here, a stochastic process is a sequence of random variables $\{X(t)\}$ ordered by a discrete time variable $t = 1, 2, 3, \ldots$

The properties of a stochastic process must generally be determined from a single time series or realization. To do this, several assumptions are usually made. First, one generally assumes that the process is stationary. This means that the probability distribution of the process is not changing over time. In addition, if a process is strictly stationary, the joint distribution of the random variables $X(t_1), \ldots, X(t_n)$ is identical to the joint distribution of $X(t_1 + t), \ldots, X(t_n + t)$ for any $t$; the joint distribution depends only on the differences $t_i - t_j$ between the times of occurrence of the events.

For a stationary stochastic process, one can write the mean and variance as

$$\mu_X = \mathrm{E}[X(t)] \tag{7.108}$$

and

$$\sigma^2 = \mathrm{Var}[X(t)] \tag{7.109}$$

Both are independent of time $t$. The *autocorrelations*, the correlation of $X$ with itself, are given by

$$\rho_X(k) = \mathrm{Cov}[X(t), X(t+k)]/\sigma_X^2 \tag{7.110}$$

for any positive integer time lag $k$. These are the statistics most often used to describe stationary stochastic processes.

When one has available only a single time series, it is necessary to estimate the values of $\mu_X$, $\sigma_X^2$, and $\rho_X(k)$ from values of the random variable that one has observed. The mean and variance are generally estimated essentially as they were in Equation 7.39.

$$\hat{\mu}_X = \bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t \tag{7.111}$$

$$\sigma_X^2 = \frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})^2 \tag{7.112}$$

while the autocorrelations $\rho_X(k)$ for any time lag $k$ can be estimated as (Jenkins and Watts, 1968)

$$\hat{\rho}_X(k) = r_k = \frac{\sum_{t=1}^{T-k}(x_{t+k} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \tag{7.113}$$

The sampling distribution of these estimators depends on the correlation structure of the stochastic process giving rise to the time series. In particular, when the observations are positively correlated, as is usually the case in natural streamflows or annual benefits in a river basin simulation, the variances of the estimated $\bar{x}$ and $\hat{\sigma}_X^2$ are larger than would be the case if the observations were independent. It is sometimes wise to take this inflation into account. Section 7.3 discusses the sampling distribution of these statistics.

All of this analysis depends on the assumption of stationarity, for only then do the quantities defined in Equations 7.108 to 7.110 have the intended meaning. Stochastic processes are not always stationary. Agricultural and urban development, deforestation, climatic variability and changes in regional resource management can alter the distribution of rainfall, streamflows, pollutant concentrations, sediment loads and groundwater levels over time. If a stochastic process is not essentially stationary over the time span in question, then statistical techniques that rely on the stationary assumption do not apply and the problem generally becomes much more difficult.

## 7.2. Markov Processes and Markov Chains

A common assumption in many stochastic water resources models is that the stochastic process $X(t)$ is a *Markov process*. A first-order Markov process has the property that the dependence of future values of the process on past values depends only on the current value and not on previous values or observations. In symbols for $k > 0$,

$$F_X[X(t+k) \mid X(t), X(t-1), X(t-2), \ldots] = F_X[X(t+k) \mid X(t)] \quad (7.114)$$

For Markov processes, the current value summarizes the state of the processes. As a consequence, the current value of the process is often referred to as the *state*. This makes physical sense as well when one refers to the state or level of an aquifer or reservoir.

A special kind of Markov process is one whose state $X(t)$ can take on only discrete values. Such a process is called a *Markov chain*. Often in water resources planning, continuous stochastic processes are approximated by Markov chains. This is done to facilitate the construction of simple stochastic models. This section presents the basic notation and properties of Markov chains.

Consider a stream whose annual flow is to be represented by a discrete random variable. Assume that the distribution of streamflows is stationary. In the following development, the continuous random variable representing the annual streamflows (or some other process) is approximated by a random variable $Q_y$ in year $y$, which takes on only $n$ discrete values $q_i$ (each value representing a continuous range or interval of possible streamflows) with unconditional probabilities $p_i$ where

$$\sum_{i=1}^{n} p_i = 1 \quad (7.115)$$

It is frequently the case that the value of $Q_{y+1}$ is not independent of $Q_y$. A Markov chain can model such dependence. This requires specification of the *transition probabilities* $p_{ij}$,

$$p_{ij} = \Pr[Q_{y+1} = q_j \mid Q_y = q_i] \quad (7.116)$$

A transition probability is the conditional probability that the next state is $q_j$, given that the current state is $q_i$. The transition probabilities must satisfy

$$\sum_{j=1}^{n} p_{ij} = 1 \quad \text{for all } i \quad (7.117)$$

Figure 7.11 shows a possible set of transition probabilities in a matrix. Each element $p_{ij}$ in the matrix is the probability of a transition from streamflow $q_i$ in one year to streamflow $q_j$ in the next. In this example, a low flow tends to be followed by a low flow, rather than a high flow, and vice versa.

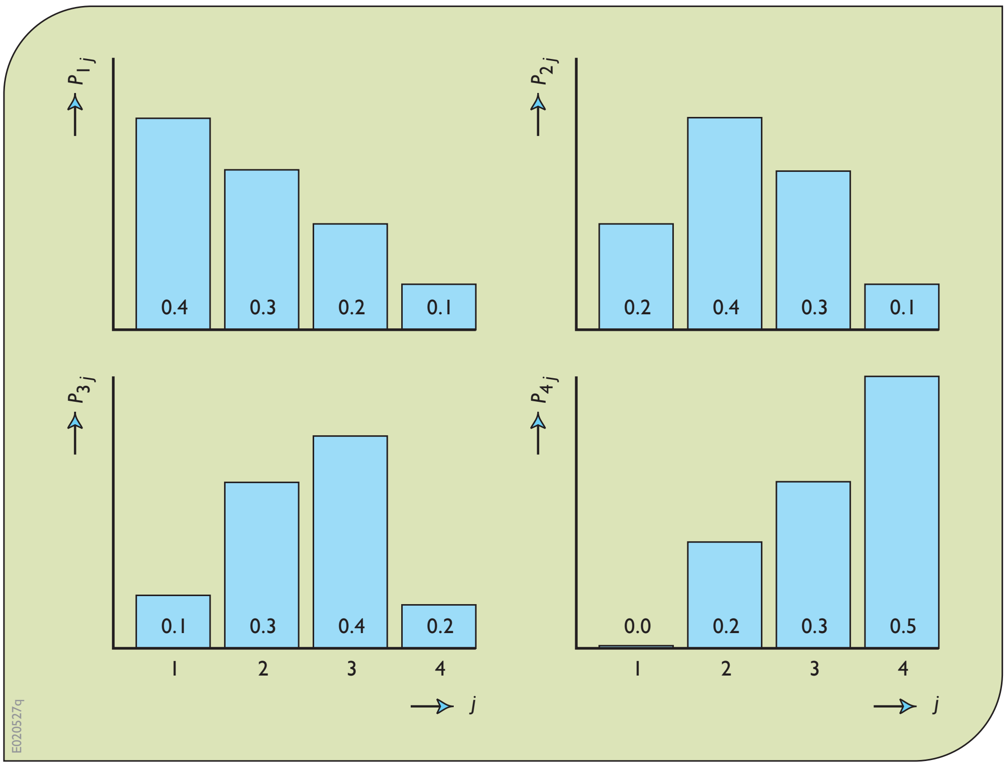



**Figure 7.11.** Matrix (above) and histograms (below) of streamflow transition probabilities showing probability of streamflow $q_j$ (represented by index $j$) in year $y+1$ given streamflow $q_i$ (represented by index $i$) in year $y$.

Let $\boldsymbol{P}$ be the transition matrix whose elements are $p_{ij}$. For a Markov chain, the transition matrix contains all the information necessary to describe the behaviour of the process. Let $p_i^y$ be the probability that the streamflow $Q_y$ is $q_i$ (in state $i$) in year $y$. Then the probability that $Q_{y+1} = q_j$ is the sum of the products of the probabilities $p_i^y$ that $Q_y = q_i$ times the probability $p_{ij}$ that the next state $Q_{y+1}$ is $q_j$ given that $Q_y = q_i$. In symbols, this relationship is written:

$$p_j^{y+1} = p_1^y p_{1j} + p_2^y p_{2j} + \cdots + p_n^y p_{nj} = \sum_{i=1}^{n} p_i^y p_{ij} \tag{7.118}$$

Letting $\boldsymbol{p}^y$ be the row vector of state resident probabilities $(p_i^y, \ldots, p_n^y)$, this relationship may be written

$$\boldsymbol{p}^{(y+1)} = \boldsymbol{p}^{(y)}\boldsymbol{P} \tag{7.119}$$

To calculate the probabilities of each streamflow state in year $y + 2$, one can use $\boldsymbol{p}^{(y+1)}$ in Equation 7.119 to obtain

$$\boldsymbol{p}^{(y+2)} = \boldsymbol{p}^{(y+1)}\boldsymbol{P} \text{ or } \boldsymbol{p}^{(y+2)} = \boldsymbol{p}^y\boldsymbol{P}^2$$

Continuing in this manner, it is possible to compute the probabilities of each possible streamflow state for years $y + 1, y + 2, y + 3, \ldots, y + k, \ldots$ as

$$\boldsymbol{p}^{(y+k)} = \boldsymbol{p}^y\boldsymbol{P}^k \tag{7.120}$$

Returning to the four-state example in Figure 7.11, assume that the flow in year $y$ is in the interval represented by $q_2$. Hence in year $y$ the unconditional streamflow probabilities $p_i^y$ are (0, 1, 0, 0). Knowing each $p_i^y$, the probabilities $p_j^{y+1}$corresponding to each of the four streamflow states can be determined. From Figure 7.11, the probabilities $p_j^{y+1}$ are 0.2, 0.4, 0.3 and 0.1 for $j = 1, 2, 3$ and 4, respectively. The probability vectors for nine future years are listed in Table 7.8.

As time progresses, the probabilities generally reach limiting values. These are the *unconditional* or *steady-state* probabilities. The quantity $p_i$ has been defined as the unconditional probability of $q_i$. These are the steady-state probabilities which $\boldsymbol{p}^{(y+k)}$ approaches for large $k$. It is clear from Table 7.8 that as $k$ becomes larger, Equation 7.118 becomes

$$p_j = \sum_{i=1}^{n} p_i p_{ij} \tag{7.121}$$

or in vector notation, Equation 7.119 becomes

$$\boldsymbol{p} = \boldsymbol{p}\boldsymbol{P} \tag{7.122}$$

| year | $P_1^y$ | $P_2^y$ | $P_3^y$ | $P_4^y$ |
|---|---|---|---|---|
| $y$ | 0.000 | 1.000 | 0.000 | 0.000 |
| $y + 1$ | 0.200 | 0.400 | 0.300 | 0.100 |
| $y + 2$ | 0.190 | 0.330 | 0.310 | 0.170 |
| $y + 3$ | 0.173 | 0.316 | 0.312 | 0.199 |
| $y + 4$ | 0.163 | 0.312 | 0.314 | 0.211 |
| $y + 5$ | 0.159 | 0.310 | 0.315 | 0.216 |
| $y + 6$ | 0.157 | 0.309 | 0.316 | 0.218 |
| $y + 7$ | 0.156 | 0.309 | 0.316 | 0.219 |
| $y + 8$ | 0.156 | 0.309 | 0.316 | 0.219 |
| $y + 9$ | 0.156 | 0.309 | 0.316 | 0.219 |

E021101h

**Table 7.8.** Successive streamflow probabilities based on transition probabilities in Figure 7.11.

where $\boldsymbol{p}$ is the row vector of unconditional probabilities $(p_1, \ldots, p_n)$. For the example in Table 7.8, the probability vector $\boldsymbol{p}$ equals (0.156, 0.309, 0.316, 0.219).

The steady-state probabilities for any Markov chain can be found by solving simultaneous Equation 7.122 for all but one of the states $j$ together with the constraint

$$\sum_{i=1}^{n} p_i = 1 \tag{7.123}$$

Annual streamflows are seldom as highly correlated as the flows in this example. However, monthly, weekly and especially daily streamflows generally have high serial correlations. Assuming that the unconditional steady-state probability distributions for monthly streamflows are stationary, a Markov chain can be defined for each month's streamflow. Since there are twelve months in a year, there would be twelve transition matrices, the elements of which could be denoted as $p_{ij}^t$. Each defines the probability of a streamflow $q_j^{t+1}$ in month $t + 1$, given a streamflow $q_i^t$ in month $t$. The steady-state stationary probability vectors for each month can be found by the procedure outlined above, except that now all twelve matrices are used to calculate all twelve steady-state probability vectors. However, once the steady-state vector $\boldsymbol{p}$ is found for one month, the others are easily computed using Equation 7.120 with $t$ replacing $y$.

## 7.3. Properties of Time-Series Statistics

The statistics most frequently used to describe the distribution of a continuous-state stationary stochastic process are the sample mean, variance and various autocorrelations. Statistical dependence among the observations, as is frequently found in time series, can have a marked effect on the distribution of these statistics. This part of Section 7 reviews the sampling properties of these statistics when the observations are a realization of a stochastic process.

The sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{7.124}$$

when viewed as a random variable is an unbiased estimate of the mean of the process $\mu_X$, because

$$E[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu_X \tag{7.125}$$

However, correlation among the $X_i$'s, so that $\rho_X(k) \neq 0$ for $k > 0$, affects the variance of the estimated mean $\bar{X}$.

$$\begin{aligned}\mathrm{Var}(\bar{X}) &= E[(\bar{X} - \mu_X)^2] \\ &= \frac{1}{n^2} E\left\{\sum_{t=1}^{n}\sum_{s=1}^{n}(X_t - \mu_X)(X_s - \mu_X)\right\} \\ &= \frac{\sigma_X^2}{n}\left\{1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_X(k)\right\}\end{aligned} \tag{7.126}$$

The variance of $\bar{X}$, equal to $\sigma_x^2/n$ for independent observations, is inflated by the factor within the brackets. For $\rho_X(k) \geq 0$, as is often the case, this factor is a nondecreasing function of $n$, so that the variance of $\bar{X}$ is inflated by a factor whose importance does not decrease with increasing sample size. This is an important observation, because it means that the average of a correlated time series will be less precise than the average of a sequence of independent random variables of the same length with the same variance.

A common model of stochastic series has

$$\rho_X(k) = [\rho_X(1)]^k = \rho^k \tag{7.127}$$

This correlation structure arises from the autoregressive Markov model discussed at length in Section 8. For this correlation structure

| sample size $n$ | correlation of consecutive observations $\rho = 0.0$ | $\rho = 0.3$ | $\rho = 0.6$ |
|---|---|---|---|
| 25 | 0.050 | 0.067 | 0.096 |
| 50 | 0.035 | 0.048 | 0.069 |
| 100 | 0.025 | 0.034 | 0.050 |

E021101j

Table 7.9. Standard error of $\bar{X}$ when $\sigma_x = 0.25$ and $\rho_X(k) = \rho^k$.

$$\mathrm{Var}(\bar{X}) = \frac{\sigma_X^2}{n}\left\{1 + \frac{2\rho}{n}\frac{[n(1-\rho) - (1-\rho^n)]}{(1-\rho)^2}\right\} \tag{7.128}$$

Substitution of the sample estimates for $\sigma_X^2$ and $\rho_X(1)$ in the equation above often yields a more realistic estimate of the variance of $\bar{X}$ than does the estimate $s_X^2/n$ if the correlation structure $\rho_X(k) = \rho^k$ is reasonable; otherwise, Equation 7.126 may be employed. Table 7.9 illustrates the effect of correlation among the $X_t$ values on the standard error of their mean, equal to the square root of the variance in Equation 7.126.

The properties of the estimate of the variance of $X$,

$$\hat{\sigma}_X^2 = v_X^2 = \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2 \tag{7.129}$$

are also affected by correlation among the $X_t$'s. Here $v$ rather than $s$ is used to denote the variance estimator, because $n$ is employed in the denominator rather than $n - 1$. The expected value of $v_x^2$ becomes

$$E[v_X^2] = \sigma_X^2\left\{1 - \frac{1}{n} - \frac{2}{n}\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_X(k)\right\} \tag{7.130}$$

The bias in $v_X^2$ depends on terms involving $\rho_X(1)$ through $\rho_X(n-1)$. Fortunately, the bias in $v_X^2$ decreases with $n$ and is generally unimportant when compared to its variance.

Correlation among the $X_t$'s also affects the variance of $v_X^2$. Assuming that $X$ has a normal distribution (here the variance of $v_X^2$ depends on the fourth moment of $X$), the variance of $v_X^2$ for large $n$ is approximately (Kendall and Stuart, 1966, Sec. 48.1).

$$\mathrm{Var}(v_X^2) \cong 2\frac{\sigma_X^4}{n}\left\{1 + 2\sum_{k=1}^{\infty}\rho_X^2(k)\right\} \tag{7.131}$$

where for $\rho_X(k) = \rho^k$, Equation 7.131 becomes

$$\mathrm{Var}(v_X^2) \cong 2\frac{\sigma_X^4}{n}\left(\frac{1+\rho^2}{1-\rho^2}\right) \quad (7.132)$$

Like the variance of $\bar{X}$, the variance of $v_X^2$ is inflated by a factor whose importance does not decrease with $n$. This is illustrated by Table 7.10 which gives the standard deviation of $v_X^2$ divided by the true variance $\sigma_X^2$ as a function of $n$ and $\rho$ when the observations have a normal distribution and $\rho_X(k) = \rho^k$. This would be the coefficient of variation of $v_X^2$ were it not biased.

A fundamental problem of time-series analyses is the estimation or description of the relationship among the random variable values at different times. The statistics used to describe this relationship are the autocorrelations. Several estimates of the autocorrelations have been suggested. A simple and satisfactory estimate recommended by Jenkins and Watts (1968) is:

$$\hat{\rho}_X(k) = r_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \quad (7.133)$$

Here $r_k$ is the ratio of two sums where the numerator contains $n - k$ terms and the denominator contains $n$ terms. The estimate $r_k$ is biased, but unbiased estimates frequently have larger mean square errors (Jenkins and Watts, 1968). A comparison of the bias and variance of $r_1$ is provided by the case when the $X_t$'s are independent normal variates. Then (Kendall and Stuart, 1966)

$$E[r_1] = -\frac{1}{n} \quad (7.134a)$$

and

$$\mathrm{Var}(r_1) = \frac{(n-2)^2}{n^2(n-1)} \cong \frac{1}{n} \quad (7.134b)$$

For $n = 25$, the expected value of $r_1$ is $-0.04$ rather than the true value of zero; its standard deviation is 0.19. This results in a mean square error of $(E[r_1])^2 + \mathrm{Var}(r_1) = 0.0016 + 0.0353 = 0.0369$. Clearly, the variance of $r_1$ is the dominant term.

For $X_t$ values that are not independent, exact expressions for the variance of $r_k$ generally are not available.

| sample size $n$ | correlation of consecutive observations $\rho = 0.0$ | $\rho = 0.3$ | $\rho = 0.6$ |
|---|---|---|---|
| 25 | 0.28 | 0.31 | 0.41 |
| 50 | 0.20 | 0.22 | 0.29 |
| 100 | 0.14 | 0.15 | 0.21 |

E021101k

**Table 7.10.** Standard deviation of $[v_X^2/\sigma_X^2]$ when observations have a normal distribution and $\rho_X(k) = \rho^k$.

However, for normally distributed $X_t$ and large $n$ (Kendall and Stuart, 1966),

$$\mathrm{Var}(r_k) \cong \frac{1}{n}\sum_{l=-\infty}^{+\infty}\left[\rho_X^2(l) + \rho_X(l+k)\rho_X(l-k) - 4\rho_X(k)\rho_X(l)\rho_X(k-l) + 2\rho_X^2(k)\rho_X^2(l)\right] \quad (7.135)$$

If $\rho_X(k)$ is essentially zero for $k > q$, then the simpler expression (Box et al., 1994)

$$\mathrm{Var}(r_k) \cong \frac{1}{n}\left[1 + 2\sum_{t=1}^{q}\rho_X^2(l)\right] \quad (7.136)$$

is valid for $r_k$ corresponding to $k > q$. Thus for large $n$, $\mathrm{Var}(r_k) \geq l/n$ and values of $r_k$ will frequently be outside the range of $\pm 1.65/\sqrt{n}$, even though $\rho_X(k)$ may be zero.

If $\rho_X(k) = \rho^k$, Equation 7.136 reduces to

$$\mathrm{Var}(r_k) \cong \frac{1}{n}\left[\frac{(l+\rho^2)(l-\rho^{2k})}{l-\rho^2} - 2k\rho^{2k}\right] \quad (7.137)$$

In particular for $r_1$, this gives

$$\mathrm{Var}(r_1) \cong \frac{1}{n}(1-\rho^2) \quad (7.138)$$

Approximate values of the standard deviation of $r_1$ for different values of $n$ and $\rho$ are given in Table 7.11.

The estimates of $r_k$ and $r_{k+j}$ are highly correlated for small $j$; this causes plots of $r_k$ versus $k$ to exhibit slowly varying cycles when the true values of $\rho_X(k)$ may be zero. This increases the difficulty of interpreting the sample autocorrelations.

| sample size $n$ | correlation of consecutive observations $\rho = 0.0$ | $\rho = 0.3$ | $\rho = 0.6$ |
|---|---|---|---|
| 25 | 0.20 | 0.19 | 0.16 |
| 50 | 0.14 | 0.13 | 0.11 |
| 100 | 0.10 | 0.095 | 0.080 |

E021101m

**Table 7.11.** Approximate standard deviation of $r_1$ when observations have a normal distribution and $\rho_X(k) = \rho^k$.

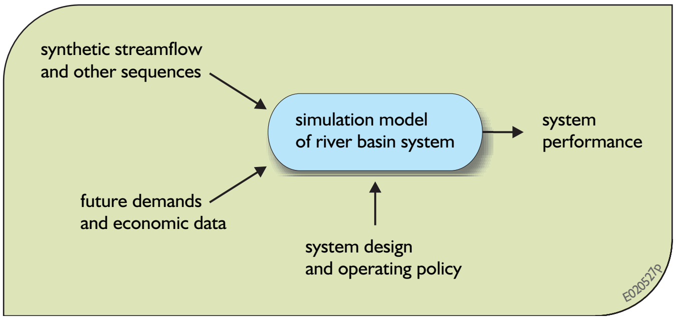



**Figure 7.12.** Structure of a simulation study, indicating the transformation of a synthetic streamflow sequence, future demands and a system design and operating policy into system performance statistics.

# 8. Synthetic Streamflow Generation

## 8.1. Introduction

This section is concerned primarily with ways of generating sample data such as streamflows, temperatures and rainfall that are used in water resource systems simulation studies (e.g., as introduced in the next section). The models and techniques discussed in this section can be used to generate any number of quantities used as inputs to simulation studies. For example Wilks (1998, 2002) discusses the generation of wet and dry days, rainfall depths on wet days and associated daily temperatures. The discussion here is directed toward the generation of streamflows because of the historical development and frequent use of these models in that context (Matalas and Wallis, 1976). In addition, they are relatively simple compared to more complete daily weather generators and many other applications. Generated streamflows have been called *synthetic* to distinguish them from historical observations (Fiering, 1967). The activity has been called stochastic hydrological modelling. More detailed presentations can be found in Marco et al. (1989) and Salas (1993).

River basin simulation studies can use many sets of streamflow, rainfall, evaporation and/or temperature sequences to evaluate the statistical properties of the performance of alternative water resources systems. For this purpose, synthetic flows and other generated quantities should resemble, statistically, those sequences that are likely to be experienced during the planning period. Figure 7.12 illustrates how synthetic streamflow, rainfall and other stochastic sequences are used in conjunction with projections of future demands and other economic data to determine how different system designs and operating policies might perform.

Use of only the historical flow or rainfall record in water resource studies does not allow for the testing of alternative designs and policies against the range of sequences that are likely to occur in the future. We can be very confident that the future historical sequence of flows will not be the historical one, yet there is important information in that historical record. That information is not fully used if only the historical sequence is simulated. Fitting continuous distributions to the set of historical flows and then using those distributions to generate other sequences of flows, all of which are statistically similar and equally weighted, gives one a broader range of inputs to simulation models. Testing designs and policies against that broader range of flow sequences that could occur more clearly identifies the variability and range of possible future performance indicator values. This in turn should lead to the selection of more robust system designs and policies.

The use of synthetic streamflows is particularly useful for water resources systems having large amounts of over-year storage. Use of only the historical hydrological record in system simulation yields only one time history of how the system would operate from year to year. In water resources systems with relatively little storage, so that reservoirs and/or groundwater aquifers refill almost every year, synthetic hydrological sequences may not be needed if historical sequences of a reasonable length are available. In this second case, a twenty-five-year historical record provides twenty-five descriptions of the possible within-year operation of the system. This may be sufficient for many studies.

Generally, use of stochastic sequences is thought to improve the precision with which water resources system performance indices can be estimated, and some studies have shown this to be the case (Vogel and Shallcross, 1996; Vogel and Stedinger, 1988). In particular, if system operation performance indices have thresholds and sharp breaks, then the coarse descriptions provided by historical series are likely to provide relative inaccurate estimates of the expected values of such statistics. For example, suppose that shortages only invoke non-linear penalties on average one year in twenty. Then in a sixty-year simulation there is a 19% probability that the penalty will be invoked at most once, and an 18% probability it will be invoked five or more times. Thus the calculation of the annual average value of the penalty would be highly unreliable unless some smoothing of the input distributions is allowed, associated with a long simulation analysis.

On the other hand, if one is only interested in the mean flow, or average benefits that are mostly a linear function of flows, then use of stochastic sequences will probably add little information to what is obtained simply by simulating the historical record. After all, the fitted models are ultimately based on the information provided in the historical record, and their use does not produce new information about the hydrology of the basin.

If in a general sense one has available *n* years of record, the statistics of that record can be used to build a stochastic model for generating thousands of years of flow. These synthetic data can then be used to estimate more exactly the system performance, assuming, of course, that the flow-generating model accurately represents nature. But the initial uncertainty in the model parameters resulting from having only *n* years of record would still remain (Schaake and Vicens, 1980).

An alternative is to run the historical record (if it is sufficiently complete at every site and contains no gaps of missing data) through the simulation model to generate *n* years of output. That output series can be processed to produce estimates of system performance. So the question is the following: Is it better to generate multiple input series based on uncertain parameter values and use those to determine average system performance with great precision, or is it sufficient just to model the *n*-year output series that results from simulation of the historical series?

The answer seems to depend upon how well behaved the input and output series are. If the simulation model is linear, it does not make much difference. If the simulation model were highly non-linear, then modelling the input series would appear to be advisable. Or if one is developing reservoir operating policies, there is a tendency to make a policy sufficiently complex to deal very well with the few droughts in the historical record, giving a false sense of security and likely misrepresenting the probability of system performance failures.

Another situation where stochastic data generating models are useful is when one wants to understand the impact on system performance estimates of the parameter uncertainty stemming from short historical records. In that case, parameter uncertainty can be incorporated into streamflow generating models so that the generated sequences reflect both the variability that one would expect in flows over time as well as the uncertainty of the parameter values of the models that describe that variability (Valdes et al., 1977; Stedinger and Taylor, 1982a,b; Stedinger et al., 1985; Vogel and Stedinger, 1988).

If one decides to use a stochastic data generator, the challenge is to use a model that appropriately describes the important relationships, but does not attempt to reproduce more relationships than are justified or can be estimated with available data sets.

Two basic techniques are used for streamflow generation. If the streamflow population can be described by a stationary stochastic process (a process whose parameters do not change over time), and if a long historical streamflow record exists, then a stationary stochastic streamflow model may be fit to the historical flows. This statistical model can then generate synthetic sequences that describe selected characteristics of the historical flows. Several such models are discussed below.

The assumption of stationarity is not always plausible, particularly in river basins that have experienced marked changes in runoff characteristics due to changes in land cover, land use, climate or the use of groundwater during the period of flow record. Similarly, if the physical characteristics of a basin change substantially in the future, the historical streamflow record may not provide reliable estimates of the distribution of future unregulated flows. In the absence of the stationarity of streamflows or a representative historical record, an alternative scheme is to assume that precipitation is a stationary stochastic process and to route either historical or synthetic precipitation sequences through an appropriate rainfall–runoff model of the river basin.

## 8.2. Streamflow Generation Models

The first step in the construction of a statistical streamflow generating model is to extract from the historical streamflow record the fundamental information about the joint distribution of flows at different sites and at different times. A streamflow model should ideally capture what is judged to be the fundamental characteristics of the joint distribution of the flows. The specification of what characteristics are fundamental is of primary importance.

One may want to model as closely as possible the true marginal distribution of seasonal flows and/or the marginal distribution of annual flows. These describe both how much water may be available at different times and also how variable is that water supply. Also, modelling the joint distribution of flows at a single site in different months, seasons and years may be appropriate. The persistence of high flows and of low flows, often described by their correlation, affects the reliability with which a reservoir of a given size can provide a given yield (Fiering, 1967; Lettenmaier and Burges, 1977a, 1977b; Thyer and Kuczera, 2000). For multi-component reservoir systems, reproduction of the joint distribution of flows at different sites and at different times will also be important.

Sometimes, a streamflow model is said to resemble statistically the historical flows if it produces flows with the same mean, variance, skew coefficient, autocorrelations and/or cross-correlations as were observed in the historical series. This definition of statistical resemblance is attractive because it is operational and only requires an analyst to only find a model that can reproduce the observed statistics. The drawback of this approach is that it shifts the modelling emphasis away from trying to find a good model of marginal distributions of the observed flows and their joint distribution over time and over space, given the available data, to just reproducing arbitrarily selected statistics. Defining statistical resemblance in terms of moments may also be faulted for specifying that the parameters of the fitted model should be determined using the observed sample moments, or their unbiased counterparts. Other parameter estimation techniques, such as maximum likelihood estimators, are often more efficient.

Definition of resemblance in terms of moments can also lead to confusion over whether the population parameters should equal the sample moments, or whether the fitted model should generate flow sequences whose sample moments equal the historical values. The two concepts are different because of the biases (as discussed in Section 7) in many of the estimators of variances and correlations (Matalas and Wallis, 1976; Stedinger, 1980, 1981; Stedinger and Taylor, 1982a).

For any particular river basin study, one must determine what streamflow characteristics need to be modelled. The decision should depend on what characteristics are important to the operation of the system being studied, the available data, and how much time can be spared to build and test a stochastic model. If time permits, it is good practice to see if the simulation results are in fact sensitive to the generation model and its parameter values by using an alternative model and set of parameter values. If the model's results are sensitive to changes, then, as always, one must exercise judgement in selecting the appropriate model and parameter values to use.

This section presents a range of statistical models for the generation of synthetic data. The necessary sophistication of a data-generating model depends on the intended use of the generated data. Section 8.3 below presents the simple autoregressive Markov model for generating annual flow sequences. This model alone is too simple for many practical studies, but is useful for illustrating the fundamentals of the more complex models that follow. It seems, therefore, worth some time exploring the properties of this basic model.

Subsequent sections discuss how flows with any marginal distribution can be produced, and present models for generating sequences of flows that can reproduce the persistence of historical flow sequences. Other parts of this section present models for generating concurrent flows at several sites and for generating seasonal or monthly flows that preserve the characteristics of annual flows. More detailed discussions for those wishing to study synthetic streamflow models in greater depth can be found in Marco et al. (1989) and Salas (1993).

## 8.3. A Simple Autoregressive Model

A simple model of annual streamflows is the autoregressive Markov model. The historical annual flows $q_y$ are thought of as particular values of a stationary stochastic process $Q_y$. The generation of annual streamflows and other variables would be a simple matter if annual flows were independently distributed. In general, this is not the case and a generating model for many phenomena should capture the relationship between values in different years or in other time periods. A common and reasonable assumption is that annual flows are the result of a first-order Markov process (as discussed in Section 7.2).

Assume for now that annual streamflows are normally distributed. In some areas the distribution of annual flows is in fact nearly normal. Streamflow models that produce non-normal streamflows are discussed as an extension of this simple model.

The joint normal density function of two streamflows $Q_y$ and $Q_w$ in years $y$ and $w$ having mean $\mu$, variance $\sigma^2$, and year-to-year correlation $\rho$ between flows is

$$f(q_y, q_w) = \frac{1}{2\pi\sigma^2(1-\rho^2)^{0.5}} \exp\left[\frac{(q_y-\mu)^2 - 2\rho(q_y-\mu)(q_w-\mu) + (q_w-\mu)^2}{2\sigma^2(1-\rho^2)}\right] \tag{7.139}$$

The joint normal distribution for two random variables with the same mean and variance depend only on their common mean $\mu$, variance $\sigma^2$, and the correlation $\rho$ between the two (or equivalently the covariance $\rho\sigma^2$).

The sequential generation of synthetic streamflows requires the conditional distribution of the flow in one year given the value of the flows in previous years. However, if the streamflows are a first-order (lag 1) Markov process, then the distribution of the flow in year $y + 1$ depends entirely on the value of the flow in year $y$. In addition, if the annual streamflows have a multivariate normal distribution, then the conditional distribution of $Q_{y+1}$ is normal with mean and variance

$$E[Q_{y+1} \mid Q_y = q_y] = \mu + \rho(q_y - \mu) \tag{7.140}$$

$$\text{Var}(Q_{y+1} \mid Q_y = q_y) = \sigma^2(1 - \rho^2)$$

where $q_y$ is the value of the random variable $Q_y$ in year $y$. This relationship is illustrated in Figure 7.13. Notice that the larger the absolute value of the correlation $\rho$ between the flows, the smaller the conditional variance of $Q_{y+1}$, which in this case does not depend at all on the value $q_y$.

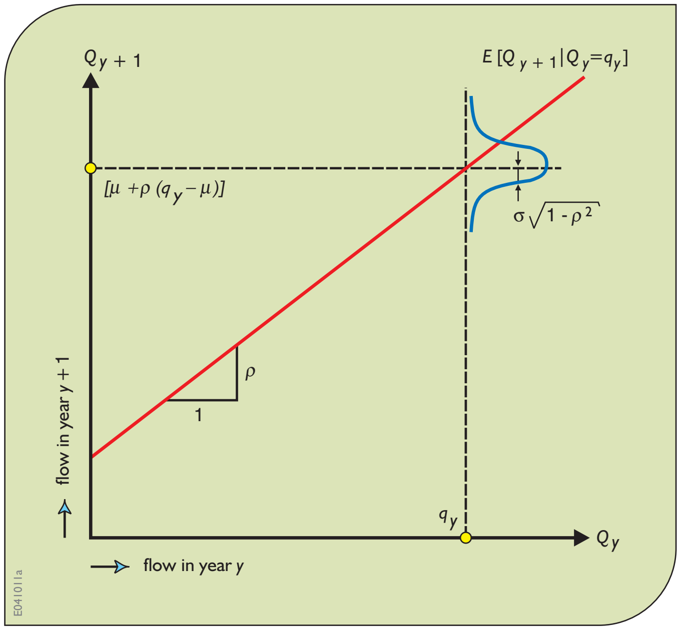



**Figure 7.13. Conditional distribution of $Q_{y+1}$ given $Q_y = q_y$ for two normal random variables.**

Synthetic normally distributed streamflows that have mean $\mu$, variance $\sigma^2$, and year-to-year correlation $\rho$, are produced by the model

$$Q_{y+1} = \mu + \rho(Q_y - \mu) + V_y\sigma\sqrt{1-\rho^2} \tag{7.141}$$

where $V_y$ is a standard normal random variable, meaning that it has zero mean, $E[V_y] = 0$, and unit variance, $E\left[V_y^2\right] = 1$. The random variable $V_y$ is added here to provide the variability in $Q_{y+1}$ that remains even after $Q_y$ is known. By construction, each $V_y$ is independent of past flows $Q_w$ where $w \leq y$, and $V_y$ is independent of $V_w$ for $w \neq y$. These restrictions imply that

$$E[V_w V_y] = 0 \quad \text{for} \quad w \neq y \tag{7.142}$$

and

$$E[(Q_w - \mu)V_y] = 0 \quad \text{for} \quad w \leq y \tag{7.143}$$

Clearly, $Q_{y+1}$ will be normally distributed if both $Q_y$ and $V_y$ are normally distributed because sums of independent normally distributed random variables are normally distributed.

It is a straightforward procedure to show that this basic model indeed produces streamflows with the specified moments, as demonstrated below.

Using the fact that $E[V_y] = 0$, the conditional mean of $Q_{y+1}$ given that $Q_y$ equals $q_y$ is

$$\begin{aligned} E[Q_{y+1} \mid q_y] &= E[\mu + \rho(q_y - \mu) + V_y \sigma\sqrt{1-\rho^2}] \\ &= \mu + \rho(q_y - \mu) \end{aligned} \quad (7.144)$$

Since $E[V_y^2] = \mathrm{Var}[V_y] = 1$, the conditional variance of $Q_{y+1}$ is

$$\begin{aligned} \mathrm{Var}[Q_{y+1} \mid q_y] &= E[\{Q_{y+1} - E[Q_{y+1} \mid q_y]\}^2 | q_y] \\ &= E[\{\mu + \rho(q_y - \mu) \\ &\quad + V_y\sigma\sqrt{1-\rho^2} - [\mu + \rho(q_y - \mu)]\}^2 \\ &= E[V_y\sigma\sqrt{1-\rho^2}\,]^2 = \sigma^2(1-\rho^2) \end{aligned} \quad (7.145)$$

Thus, this model produces flows with the correct conditional mean and variance.

To compute the unconditional mean of $Q_{y+1}$ one first takes the expectation of both sides of Equation 7.141 to obtain

$$E[Q_{y+1}] = \mu + \rho(E[Q_y] - \mu) + E[V_y]\sigma\sqrt{1-\rho^2} \quad (7.146)$$

where $E[V_y] = 0$. If the distribution of streamflows is independent of time so that for all $y$, $E[Q_{y+1}] = E[Q_y] = E[Q]$, it is clear that $(1 - \rho)\, E[Q] = (1 - \rho)\, \mu$ or

$$E[Q] = \mu \quad (7.147)$$

Alternatively, if $Q_y$ for $y = 1$ has mean $\mu$, then Equation 7.146 indicates that $Q_2$ will have mean $\mu$. Thus repeated application of the Equation 7.146 would demonstrate that all $Q_y$ for $y > 1$ have mean $\mu$.

The unconditional variance of the annual flows can be derived by squaring both sides of Equation 7.141 to obtain

$$\begin{aligned} E[(Q_{y+1} - \mu)^2] &= E[\{\rho(Q_y - \mu) + V_y\sigma\sqrt{1-\rho^2}\}^2] \\ &= \rho^2 E[(Q_y - \mu)^2] + 2\rho\sigma\sqrt{1-\rho^2} \\ &\quad \times E[(Q_y - \mu)V_y] + \sigma^2(1-\rho^2)E[V_y^2] \end{aligned} \quad (7.148)$$

Because $V_y$ is independent of $Q_y$ (Equation 7.143), the second term on the right-hand side of Equation 7.148 vanishes. Hence the unconditional variance of $Q$ satisfies

$$E[(Q_{y+1} - \mu)^2] = \rho^2 E[(Q_y - \mu)^2] + \sigma^2(1 - \rho^2) \quad (7.149)$$

Assuming that $Q_{y+1}$ and $Q_y$ have the same variance yields

$$E[(Q - \mu)^2] = \sigma^2 \quad (7.150)$$

so that the unconditional variance is $\sigma^2$, as required.

Again, if one does not want to assume that $Q_{y+1}$ and $Q_y$ have the same variance, a recursive argument can be adopted to demonstrate that if $Q_1$ has variance $\sigma^2$, then $Q_y$ for $y \geq 1$ has variance $\sigma^2$.

The covariance of consecutive flows is another important issue. After all, the whole idea of building these time-series models is to describe the year-to-year correlation of the flows. Using Equation 7.141 one can show that the covariance of consecutive flows must be

$$\begin{aligned} E[(Q_{y+1} - \mu)(Q_y - \mu)] &= E\{[\rho(Q_y - \mu) \\ &\quad + V_y\sigma\sqrt{1-\rho^2}](Q_y - \mu)\} \\ &= \rho E[(Q_y - \mu)^2] = \rho\sigma^2 \end{aligned} \quad (7.151)$$

where $E[(Q_y - \mu)V_y] = 0$ because $V_y$ and $Q_y$ are independent (Equation 7.143).

Over a longer time scale, another property of this model is that the covariance of flows in year $y$ and $y + k$ is

$$E[(Q_{y+k} - \mu)(Q_y - \mu)] = \rho^k\sigma^2 \quad (7.152)$$

This equality can be proven by induction. It has already been shown for $k = 0$ and 1. If it is true for $k = j - 1$, then

$$\begin{aligned} E[(Q_{y+j} - \mu)(Q_y - \mu)] &= E\{[\rho(Q_{y+j-1} - \mu) \\ &\quad + V_{y+j-1}\sigma\sqrt{1-\rho^2}](Q_y - \mu)\} \\ &= \rho E[(Q_y - \mu)](Q_{y+j-1} - \mu)] \\ &= \rho[\rho^{j-1}\sigma^2] = \rho^j\sigma^2 \end{aligned} \quad (7.153)$$

where $E[(Q_y - \mu)V_{y+j-1}] = 0$ for $j \geq 1$. Hence Equation 7.152 is true for any value of $k$.

It is important to note that the results in Equations 7.144 to 7.152 do not depend on the assumption that the random variables $Q_y$ and $V_y$ are normally distributed. These relationships apply to all autoregressive Markov processes of the form in Equation 7.141 regardless of the distributions of $Q_y$ and $V_y$. However, if the flow $Q_y$ in year $y = 1$ is normally distributed with mean $\mu$ and variance $\sigma^2$, and if the $V_y$ are independent normally distributed random variables with mean zero and unit variance, then the generated $Q_y$ for $y \geq 1$ will also be normally distributed with mean $\mu$ and variance $\sigma^2$. The next section considers how this and other models can be used to generate streamflows that have other than a normal distribution.

## 8.4. Reproducing the Marginal Distribution

Most models for generating stochastic processes deal directly with normally distributed random variables. Unfortunately, flows are not always adequately described by the normal distribution. In fact, streamflows and many other hydrological data cannot really be normally distributed because of the impossibility of negative values. In general, distributions of hydrological data are positively skewed, having a lower bound near zero and, for practical purposes, an unbounded right-hand tail. Thus they look like the gamma or lognormal distribution illustrated in Figures 7.3 and 7.4.

The asymmetry of a distribution is often measured by its coefficient of skewness. In some streamflow models, the skew of the random elements $V_y$ is adjusted so that the models generate flows with the desired mean, variance and skew coefficient. For the autoregressive Markov model for annual flows

$$\begin{aligned} \mathrm{E}[(Q_{y+1}-\mu)^3] &= \mathrm{E}\left[\rho(Q_y-\mu)+V_y\sigma\sqrt{1-\rho^2}\,\right]^3 \\ &= \rho^3\,\mathrm{E}[(Q_y-\mu)^3] \\ &\quad +\sigma^3(1-\rho^2)^{3/2}\,\mathrm{E}\left[V_y^3\right] \end{aligned} \tag{7.154}$$

so that

$$\gamma_Q = \frac{\mathrm{E}[(Q-\mu)^3]}{\sigma^3} = \frac{(1-\rho^2)^{3/2}}{1-\rho^3}\,\mathrm{E}\left[V_y^3\right] \tag{7.155}$$

By appropriate choice of the skew of $V_y$, $\mathrm{E}[V_y^3]$, the desired skew coefficient of the annual flows can be produced. This method has often been used to generate flows that have approximately a gamma distribution by using $V_y$'s with a gamma distribution and the required skew. The resulting approximation is not always adequate (Lettenmaier and Burges, 1977a).

The alternative and generally preferred method is to generate normal random variables and then transform these variates to streamflows with the desired marginal distribution. Common choices for the distribution of streamflows are the two-parameter and three-parameter lognormal distributions or a gamma distribution. If $Q_y$ is a lognormally distributed random variable, then

$$Q_y = \tau + \exp(X_y) \tag{7.156}$$

where $X_y$ is a normal random variable. When the lower bound $\tau$ is zero, $Q_y$ has a two-parameter lognormal distribution. Equation 7.156 transforms the normal variates $X_y$ into lognormally distributed streamflows. The transformation is easily inverted to obtain

$$X_y = \ln(Q_y - \tau) \quad \text{for} \quad Q_y > \tau \tag{7.157}$$

where $Q_y$ must be greater than its lower bound $\tau$.

The mean, variance, skewness of $X_y$ and $Q_y$ are related by the formulas (Matalas, 1967)

$$\begin{aligned} \mu_Q &= \tau + \exp\left(\mu_X + \frac{1}{2}\sigma_X^2\right) \\ \sigma_Q^2 &= \exp\left(2\mu_X + \sigma_X^2\right)\left[\exp\left(\sigma_X^2\right) - 1\right] \\ \gamma_Q &= \frac{\exp\left(3\sigma_X^2\right) - 3\exp\left(\sigma_X^2\right) + 2}{\left[\exp\left(\sigma_X^2\right) - 1\right]^{3/2}} \\ &= 3\phi + \phi^3 \quad \text{where } \phi = \left[\exp\left(\sigma_X^2\right) - 1\right]^{1/2} \end{aligned} \tag{7.158}$$

If normal variates $X_s^y$ and $X_y^u$ are used to generate lognormally distributed streamflows $Q_s^y$ and $Q_y^u$ at sites $s$ and $u$, then the lag-$k$ correlation of the $Q_y$'s, denoted $\rho_Q(k;\, s,\, u)$, is determined by the lag-$k$ correlation of the $X$ variables, denoted $\rho_X(k;\, s,\, u)$, and their variances $\sigma_x^2(s)$ and $\sigma_x^2(u)$, where

$$\rho_Q(k;s,u) = \frac{\exp[\rho_X(k;\, s,\, u)\sigma_X(s)\sigma_X(u)] - 1}{\left\{\exp\left[\sigma_X^2(s)\right] - 1\right\}^{1/2}\ \left\{\exp\left[\sigma_X^2(u)\right] - 1\right\}^{1/2}} \tag{7.159}$$

The correlations of the $X_y^s$ can be adjusted, at least in theory, to produce the observed correlations among the $Q_y^s$ variates. However, more efficient estimates of the true correlation of the $Q_y^s$ values are generally obtained by transforming the historical flows $q_y^s$ into their normal equivalent $x_y^s = \ln(q_y^s - \tau)$ and using the historical correlations of these $X_y^s$ values as estimators of $\rho_X(k;\, s,\, u)$ (Stedinger, 1981).

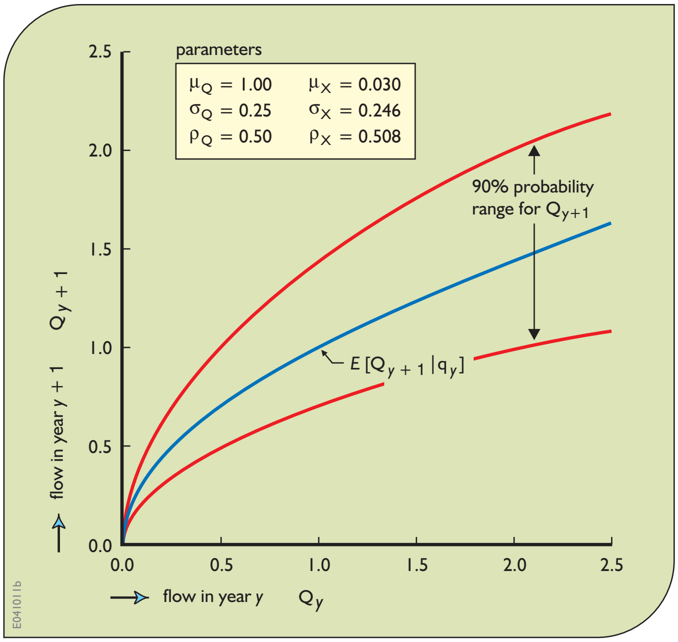



**Figure 7.14. Conditional mean of $Q_{y+1}$ given $Q_y = q_y$ and 90% probability range for the value of $Q_{y+1}$.**

Some insight into the effect of this logarithmic transformation can be gained by considering the resulting model for annual flows at a single site. If the normal variates follow the simple autoregressive Markov model

$$X_{y+1} - \mu = \rho_X(X_y - \mu) + V_y\sigma_X\sqrt{1-\rho_X^2} \tag{7.160}$$

then the corresponding $Q_y$ follow the model (Matalas, 1967)

$$Q_{y+1} = \tau + D_y\{\exp[\mu_x(1 - \rho_x)]\}(Q_y - \tau)^{\rho x} \tag{7.161}$$

where

$$D_y = \exp[(1-\rho_X^2)^{1/2}\,\sigma_X V_y] \tag{7.162}$$

The conditional mean and standard deviation of $Q_{y+1}$ given that $Q_y = q_y$ now depends on $(q_y - \tau)^{\rho x}$. Because the conditional mean of $Q_{y+1}$ is no longer a linear function of $q_y$, (as shown in Figure 7.14), the streamflows are said to exhibit differential persistence: low flows are now more likely to follow low flows than high flows are to follow high flows. This is a property often attributed to real streamflow distributions. Models can be constructed to capture the relative persistence of wet and dry periods (Matalas and Wallis, 1976; Salas, 1993; Thyer and Kuczera, 2000). Many weather generators for precipitation and temperature include such tendencies by employing a Markov chain description of the occurrence of wet and dry days (Wilks, 1998).

## 8.5. Multivariate Models

If long concurrent streamflow records can be constructed at the several sites at which synthetic streamflows are desired, then ideally a general multi-site streamflow model could be employed. O'Connell (1977), Ledolter (1978), Salas et al. (1980) and Salas (1993) discuss multivariate models and parameter estimation. Unfortunately, identification of the most appropriate model structure is very difficult for general multivariate models.

This section illustrates how the basic univariate annual-flow model in Section 8.3 can be generalized to the multivariate case. This exercise reveals how multivariate models can be constructed to reproduce specified variances and covariances of the flow vectors of interest, or some transformation of those values. This multi-site generalization of the annual AR(1) or autoregressive Markov model follows the approach taken by Matalas and Wallis (1976). This general approach can be further extended to generate multi-site/multi-season modelling procedures, as is done in Section 8.6, employing what have been called disaggregation models. However, while the size of the model matrices and vectors increases, the model is fundamentally the same from a mathematical viewpoint. Hence this section starts with the simpler case of an annual flow model.

For simplicity of presentation and clarity, vector notation is employed. Let $\mathbf{Z}_y = (Z_y^1, \ldots, Z_y^n)^T$ be the column vector of transformed zero-mean annual flows at sites $s = 1, 2, \ldots, n$, so that

$$E[Z_y^s] = 0 \tag{7.163}$$

In addition, let $\mathbf{V}_y = (V_y^1, \ldots, V_y^n)^T$ be a column vector of standard-normal random variables, where $V_y^s$ is independent of $V_w^r$ for $(r, w) \neq (s, y)$ and independent of past flows $Z_w^r$ where $y \geq w$. The assumption that the variables have zero mean implicitly suggests that the mean value has already been subtracted from all the variables. This makes the notation simpler and eliminates the need to include a constant term in the models. With all the random variables having zero mean, one can focus on reproducing the variances and covariances of the vectors included in a model.

A sequence of synthetic flows can be generated by the model

$$\mathbf{Z}_{y+1} = \mathrm{A}\mathbf{Z}_y + \mathrm{B}\mathbf{V}_y \quad (7.164)$$

where A and B are $(n \times n)$ matrices whose elements are chosen to reproduce the lag 0 and lag 1 cross-covariances of the flows at each site. The lag 0 and lag 1 covariances and cross-covariances can most economically be manipulated by use of the two matrices $S_0$ and $S_1$. The lag-zero covariance matrix, denoted $S_0$, is defined as

$$S_0 = \mathrm{E}[\mathbf{Z}_y \mathbf{Z}_y^{\mathrm{T}}] \quad (7.165)$$

and has elements

$$S_0(i, j) = \mathrm{E}[\mathbf{Z}_y^i \mathbf{Z}_y^j] \quad (7.166)$$

The lag-one covariance matrix, denoted $S_1$, is defined as

$$S_1 = \mathrm{E}[\mathbf{Z}_{y+1} \mathbf{Z}_y^{\mathrm{T}}] \quad (7.167)$$

and has elements

$$S_1(i, j) = \mathrm{E}[\mathbf{Z}_{y+1}^i \mathbf{Z}_y^j] \quad (7.168)$$

The covariances do not depend on $y$ because the streamflows are assumed to be stationary.

Matrix $S_1$ contains the lag 1 covariances and lag 1 cross-covariances. $S_0$ is symmetric because the cross-covariance $S_0(i, j)$ equals $S_0(j, i)$. In general, $S_1$ is not symmetric.

The variance–covariance equations that define the values of A and B in terms of $S_0$ and $S_1$ are obtained by manipulations of Equation 7.164. Multiplying both sides of that equation by $\mathbf{Z}_y^{\mathrm{T}}$ and taking expectations yields

$$\mathrm{E}[\mathbf{Z}_{y+1}\mathbf{Z}_y^{\mathrm{T}}] = \mathrm{E}[\mathrm{A}\mathbf{Z}_y\mathbf{Z}_y^{\mathrm{T}}] + \mathrm{E}[\mathrm{B}\mathbf{V}_y\mathbf{Z}_y^{\mathrm{T}}] \quad (7.169)$$

The second term on the right-hand side vanishes because the components of $\mathbf{Z}_y$ and $\mathbf{V}_y$ are independent. Now the first term in Equation 7.169, $\mathrm{E}[\mathrm{A}\mathbf{Z}_y \mathbf{Z}_y^{\mathrm{T}}]$, is a matrix whose $(i, j)$th element equals

$$\mathrm{E}\left[\sum_{k=1}^{n} a_{ik}\mathbf{Z}_y^k\mathbf{Z}_y^j\right] = \sum_{k=1}^{n} a_{ik}\mathrm{E}[\mathbf{Z}_y^k\mathbf{Z}_y^j] \quad (7.170)$$

The matrix with these elements is the same as the matrix $\mathrm{AE}\left[\mathbf{Z}_y \mathbf{Z}_y^{\mathrm{T}}\right]$.

Hence, A – the matrix of constants – can be pulled through the expectation operator just as is done in the scalar case where $\mathrm{E}[a\mathbf{Z}_y + b] = a\mathrm{E}[\mathbf{Z}_y] + b$ for fixed constants $a$ and $b$.

Substituting $S_0$ and $S_1$ for the appropriate expectations in Equation 7.169 yields

$$S_1 = \mathrm{A}S_0 \text{ or } \mathrm{A} = S_1 S_0^{-1} \quad (7.171)$$

A relationship to determine the matrix B is obtained by multiplying both sides of Equation 7.164 by its own transpose (this is equivalent to squaring both sides of the scalar equation $a = b$) and taking expectations to obtain

$$\begin{aligned}\mathrm{E}[\mathbf{Z}_{y+1}\mathbf{Z}_{y+1}^{\mathrm{T}}] = {} & \mathrm{E}[\mathrm{A}\mathbf{Z}_y\mathbf{Z}_y^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}] + \mathrm{E}[\mathrm{A}\mathbf{Z}_y\mathbf{V}_y^{\mathrm{T}}\mathrm{B}^{\mathrm{T}}] \\ & + \mathrm{E}[\mathrm{B}\mathbf{V}_y\mathbf{Z}_y\mathrm{A}^{\mathrm{T}}] + \mathrm{E}[\mathrm{B}\mathbf{V}_y\mathbf{V}_y^{\mathrm{T}}\mathrm{B}^{\mathrm{T}}] \quad (7.172)\end{aligned}$$

The second and third terms on the right-hand side of Equation 7.172 vanish because the components of $\mathbf{Z}_y$ and $\mathbf{V}_y$ are independent and have zero mean. $\mathrm{E}[\mathbf{V}_y\mathbf{V}_y^{\mathrm{T}}]$ equals the identity matrix because the components of $\mathbf{V}_y$ are independently distributed with unit variance. Thus

$$S_0 = \mathrm{A}S_0\mathrm{A}^{\mathrm{T}} + \mathrm{BB}^{\mathrm{T}} \quad (7.173)$$

Solving for the B matrix, one finds that it should satisfy the quadratic equation

$$\mathrm{BB}^{\mathrm{T}} = S_0 - \mathrm{A}S_0\mathrm{A}^{\mathrm{T}} = S_0 - S_1 S_0^{-1} S_1^{\mathrm{T}} \quad (7.174)$$

The last equation results from substitution of the relationship for A given in Equation 7.171 and the fact that $S_0$ is symmetric; hence, $S_0^{-1}$ is symmetric.

It should not be too surprising that the elements of B are not uniquely determined by Equation 7.174. The components of the random vector $\mathbf{V}_y$ may be combined in many ways to produce the desired covariances as long as B satisfies Equation 7.174. A lower triangular matrix that satisfies Equation 7.174 can be calculated by Cholesky decomposition (Young, 1968; Press et al., 1986).

Matalas and Wallis (1976) call Equation 7.164 the *lag-1 model*. They do not call it the Markov model because the streamflows at individual sites do not have the covariances of an autoregressive Markov process given in Equation 7.152. They suggest an alternative model for what they call the *Markov model*. It has the same structure as the lag-1 model except it does not preserve the lag-1 cross-covariances. By relaxing this requirement, they obtain a simpler model with fewer parameters that generates flows that have covariances of an autoregressive Markov process at each site. In their Markov model, the new A matrix is simply a diagonal matrix, whose diagonal elements are the lag-1 correlations of flows at each site:

$$A = \mathrm{diag}[\rho(1; i, i)] \tag{7.175}$$

where $\rho(1; i, i)$ is the lag-one correlation of flows at site $i$.

The corresponding B matrix depends on the new A matrix and $S_0$, where as before

$$BB^T = S_0 - AS_0A^T \tag{7.176}$$

The idea of fitting time-series models to each site separately and then correlating the innovations in those separate models to reproduce the cross-correlation between the series is a very general and useful modelling idea that has seen a number of applications with different time-series models (Matalas and Wallis, 1976; Stedinger et al., 1985; Camacho et al., 1985; Salas, 1993).

## 8.6. Multi-Season, Multi-Site Models

In most studies of surface water systems it is necessary to consider the variations of flows within each year. Streamflows in most areas have within-year variations, exhibiting wet and dry periods. Similarly, water demands for irrigation, municipal and industrial uses also vary, and the variations in demand are generally out of phase with the variation in within-year flows; more water is usually desired when streamflows are low, and less is desired when flows are high. This increases the stress on water delivery systems and makes it all the more important that time-series models of streamflows, precipitation and other hydrological variables correctly reproduce the seasonality of hydrological processes.

This section discusses two approaches to generating within-year flows. The first approach is based on the disaggregation of annual flows produced by an annual flow generator to seasonal flows. Thus the method allows for reproduction of both the annual and seasonal characteristics of streamflow series. The second approach generates seasonal flows in a sequential manner, as was done for the generation of annual flows. Thus the models are a direct generalization of the annual flow models already discussed.

### 8.6.1. Disaggregation Models

The disaggregation model proposed by Valencia and Schaake (1973) and extended by Mejia and Rousselle (1976) and Tao and Delleur (1976) allows for the generation of synthetic flows that reproduce statistics both at the annual and at the seasonal level. Subsequent improvements and variations are described by Stedinger and Vogel (1984), Maheepala and Perera (1996), Koutsoyiannis and Manetas (1996) and Tarboton et al. (1998).

Disaggregation models can be used for either multi-season single-site or multi-site streamflow generation. They represent a very flexible modelling framework for dealing with different time or spatial scales. Annual flows for the several sites in question or the aggregate total annual flow at several sites can be the input to the model (Grygier and Stedinger, 1988). These must be generated by another model, such as those discussed in the previous sections. These annual flows or aggregated annual flows are then disaggregated to seasonal values.

Let $\mathbf{Z}_y = (Z_y^1, \ldots, Z_y^N)^T$ be the column vector of $N$ transformed normally distributed annual or aggregate annual flows for $N$ separate sites or basins. Next, let $\mathbf{X}_y = (X_{1y}^1, \ldots, X_{Ty}^1, X_{1y}^2, \ldots, X_{Ty}^2, \ldots, X_{1y}^n, \ldots, X_{Ty}^n)^T$ be the column vector of $nT$ transformed normally distributed seasonal flows $X_{ty}^s$ for season $t$, year $y$, and site $s = 1, \ldots, n$.

Assuming that the annual and seasonal series, $Z_y^s$ and $X_{ty}^s$, have zero mean (after the appropriate transformation), the basic disaggregation model is

$$\mathbf{X}_y = A\mathbf{Z}_y + B\mathbf{V}_y \tag{7.177}$$

where $\mathbf{V}_y$ is a vector of $nT$ independent standard normal random variables, and A and B are, respectively, $nT \times N$ and $nT \times nT$ matrices. One selects values of the elements of A and B to reproduce the observed correlations among the elements of $\mathbf{X}_y$ and between the elements of $\mathbf{X}_y$ and $\mathbf{Z}_y$. Alternatively, one could attempt to reproduce the observed correlations of the untransformed flows as opposed to the transformed flows, although this is not always possible (Hoshi et al., 1978) and often produces poorer estimates of the actual correlations of the flows (Stedinger, 1981).

The values of A and B are determined using the matrices $S_{zz} = E[\mathbf{Z}_y\mathbf{Z}_y^T]$, $S_{xx} = E[\mathbf{X}_y\mathbf{X}_y^T]$, $S_{xz} = E[\mathbf{X}_y\mathbf{Z}_y^T]$, and $S_{zx} = E[\mathbf{Z}_y\mathbf{X}_y^T]$ where $S_{zz}$ was called $S_0$ earlier. Clearly, $S_{xz}^T = S_{zx}$. If $S_{xz}$ is to be reproduced, then by multiplying Equation 7.177 on the right by $\mathbf{Z}_y^T$ and taking expectations, one sees that A must satisfy

$$E[\mathbf{X}_y\mathbf{Z}_y^T] = E[A\mathbf{Z}_y\mathbf{Z}_y^T] \tag{7.178}$$

or

$$S_{xz} = AS_{zz} \tag{7.179}$$

Solving for the coefficient matrix A one obtains

$$A = S_{xz} S_{zz}^{-1} \tag{7.180}$$

To obtain an equation that determines the required values in the matrix B, one can multiply both sides of Equation 7.177 by their transpose and take expectations to obtain

$$S_{xx} = AS_{zz}A^T + BB^T \tag{7.181}$$

Thus, to reproduce the covariance matrix $S_{xx}$, the B matrix must satisfy

$$BB^T = S_{xx} - AS_{zz}A^T \tag{7.182}$$

Equations 7.180 and 7.182 for determining A and B are completely analogous to Equations 7.171 and 7.174 for the A and B matrices of the lag 1 models developed earlier. However, for the disaggregation model as formulated, $BB^T$, and hence the matrix B, can actually be singular or nearly so (Valencia and Schaake, 1973). This occurs because the real seasonal flows sum to the observed annual flows. Thus given the annual flow at a site and all but one $(T - 1)$ of the seasonal flows, the value of the unspecified seasonal flow can be determined by subtraction.

If the seasonal variables $X_{ty}^s$ correspond to non-linear transformations of the actual flows $Q_{ty}^s$, then $BB^T$ is generally sufficiently non-singular that a B matrix can be obtained by Cholesky decomposition. On the other hand, when the model is used to generate values of $X_{ty}^s$ to be transformed into synthetic flows $Q_{ty}^s$, the constraint that these seasonal flows should sum to the given value of the annual flow is lost. Thus the generated annual flows (equal to the sums of the seasonal flows) will deviate from the values that were to have been the annual flows. Some distortion of the specified distribution of the annual flows results. This small distortion can be ignored, or each year's seasonal flows can be scaled so that their sum equals the specified value of the annual flow (Grygier and Stedinger, 1988). The latter approach eliminates the distortion in the distribution of the generated annual flows by distorting the distribution of the generated seasonal flows. Koutsoyiannis and Manetas (1996) improve upon the simple scaling algorithm by including a step that rejects candidate vectors $\mathbf{X}_y$ if the required adjustment is too large, and instead generates another vector $\mathbf{X}_y$. This reduces the distortion in the monthly flows that results from the adjustment step.

The disaggregation model has substantial data requirements. When the dimension of $\mathbf{Z}_y$ is $n$ and the dimension of the generated vector $\mathbf{X}_y$ is $m$, the A matrix has $mn$ elements. The lower diagonal B matrix and the symmetric $S_{xx}$ matrix, upon which it depends, each have $m(m + 1)/2$ nonzero or non-redundant elements. For example, when disaggregating two aggregate annual flow series to monthly flows at five sites, $n = 2$ and $m = 12 \times 5 = 60$; thus, A has 120 elements while B and $S_{xx}$ each have 1,830 nonzero or non-redundant parameters. As the number of sites included in the disaggregation increases, the size of $S_{xx}$ and B increases rapidly. Very quickly the model can become overly parameterized, and there will be insufficient data to estimate all parameters (Grygier and Stedinger, 1988).

In particular, one can think of Equation 7.177 as a series of linear models generating each monthly flow $X_{ty}^k$ for $k = 1$, $t = 1, \ldots, 12$; $k = 2$, $t = 1, \ldots, 12$ up to $k = n$, $t = 1, \ldots, 12$ that reproduces the correlation of each $X_{ty}^k$ with all $n$ annual flows, $Z_y^k$, and all previously generated monthly flows. Then when one gets to the last flow in the last month, the model will be attempting to reproduce $n + (12n - 1) = 13n - 1$ annual to monthly and cross-correlations. Because the model implicitly includes a constant, this means one needs $k^* = 13n$ years of data to obtain a unique solution for this critical equation. For $n = 3$, $k^* = 39$. One could say that with a record length of forty years, there would be only one degree of freedom left in the residual model error variance described by B. That would be unsatisfactory.

When flows at many sites or in many seasons are required, the size of the disaggregation model can be reduced by disaggregation of the flows in stages. Such condensed models do not explicitly reproduce every season-to-season correlation (Lane, 1979; Stedinger and Vogel, 1984; Gryier and Stedinger, 1988; Koutsoyiannis and Manetas, 1996), Nor do they attempt to reproduce the cross-correlations among all the flow variates at the same site within a year (Lane, 1979; Stedinger et al., 1985). Contemporaneous models, like the Markov model developed earlier in Section 8.5, are models developed for individual sites whose innovation vectors $\mathbf{V}_y$ have the needed cross-correlations to reproduce the cross-correlations of the concurrent flows (Camacho et al., 1985), as was done in Equation 7.176. Grygier and

Stedinger (1991) describe how this can be done for a condensed disaggregation model without generating inconsistencies.

### 8.6.2. Aggregation Models

One can start with annual or seasonal flows, and break them down into flows in shorter periods representing months or weeks. Alternatively one can start with a model that describes the shortest time step flows. This latter approach has been referred to as aggregation to distinguish it from disaggregation.

One method for generating multi-season flow sequences is to convert the time series of seasonal flows $Q_{ty}$ into a homogeneous sequence of normally distributed zero-mean unit-variance random variables $Z_{ty}$. These can then be modelled by an extension of the annual flow generators that have already been discussed. This transformation can be accomplished by fitting a reasonable marginal distribution to the flows in each season so as to be able to convert the observed flows $q_{ty}^s$ into their transformed counterparts $Z_{ty}^s$, and vice versa. Particularly when shorter streamflow records are available, these simple approaches may yield a reasonable model of some streams for some studies. However, they implicitly assume that the standardized series is stationary, in the sense that the season-to-season correlations of the flows do not depend on the seasons in question. This assumption seems highly questionable.

This theoretical difficulty with the standardized series can be overcome by introducing a separate streamflow model for each month. For example, the classic Thomas–Fiering model (Thomas and Fiering, 1970) of monthly flows may be written

$$Z_{t+1,y} = \beta_t Z_{ty} + \sqrt{1 - \beta_t^2 V_{ty}} \tag{7.183}$$

where the $Z_{ty}$'s are standard normal random variables corresponding to the streamflow in season $t$ of year $y$, $\beta_t$ is the season-to-season correlation of the standardized flows, and $V_{ty}$ are independent standard normal random variables. The problem with this model is that it often fails to reproduce the correlation among non-consecutive months during a year and thus misrepresents the risk of multi-month and multi-year droughts (Hoshi et al., 1978).

For an aggregation approach to be attractive, it is necessary to use a model with greater persisitence than the Thomas–Fiering model. A general class of time-series models that allow reproduction of different correlation structures are the Box–Jenkins Autoregressive-Moving average models (Box et al., 1994). These models are presented by the notation ARMA($p$,$q$) for a model which depends on $p$ previous flows, and $q$ extra innovations $V_{ty}$. For example, Equation 7.141 would be called an AR(1) or AR(1,0) model. A simple ARMA(1,1) model is

$$Z_{t+1} = \phi_1 \cdot Z_t + V_{t+1} - \theta_1 \cdot V_t \tag{7.184}$$

The correlations of this model have the values

$$\rho_1 = (1 - \theta_1\phi_1)(\phi_1 - \theta_1)/(1 + \theta_1^2 - 2\phi_1\theta_1) \tag{7.185}$$

for the first lag. For $i > 1$

$$\rho_i = \phi^{i-1}\rho_1 \tag{7.186}$$

For $\phi$ values near and $0 < \theta_1 < \phi_1$, the autocorrelations $\rho_k$ can decay much slower than those of the standard AR(1) model.

The correlation function $\rho_k$ of general ARMA($p$,$q$) model,

$$Z_{t+1} = \sum_{i=1}^{p} \phi_i \cdot Z_{t+1-i} + V_{t+1} - \sum_{j=1}^{q} \theta_j \cdot V_{t+1-j} \tag{7.187}$$

is a complex function that must satisfy a number of conditions to ensure the resultant model is stationary and invertible (Box et al., 1994).

ARMA($p$,$q$) models have been extended to describe seasonal flow series by having their coefficients depend upon the season – these are called periodic autoregressive-moving average models, or PARMA. Salas and Obeysekera (1992), Salas and Fernandez (1993), and Claps et al., (1993) discuss the conceptual basis of such stochastic streamflow models. For example, Salas and Obeysekera (1992) found that low-order PARMA models, such as a PARMA(2,1), arise from reasonable conceptual representations of persistence in rainfall, runoff and groundwater recharge and release. Claps et al. (1993) observe that the PARMA(2,2) model, which may be needed if one wants to preserve year-to-year correlation, poses a parameter estimation challenge that is almost unmanageable (see also Rasmussen et al., 1996). The PARMA (1,1) model is more practical and easy to extend to the multivariate case (Hirsch, 1979; Stedinger et al., 1985; Salas, 1993; Rasmussen et al., 1996). Experience has shown that PARMA(1,1) models do a better job of reproducing the correlation of seasonal flows beyond lag 1 than does a Thomas–Fiering PAR(1,0) model (see for example, Bartolini and Salas, 1993).

# 9. Stochastic Simulation

This section introduces stochastic simulation. Much more detail on simulation is contained in later chapters. As discussed in Chapter 3, simulation is a flexible and widely used tool for the analysis of complex water resources systems. Simulation is trial and error. One must define the system being simulated, both its design and operating policy, and then simulate it to see how it works. If the purpose is to find the best design and operating policy, many such alternatives must be simulated and their results must be compared. When the number of alternatives to simulate becomes too large for the time and money available for such analyses, some kind of preliminary screening, perhaps using optimization models, may be justified. This use of optimization for preliminary screening – that is, for eliminating alternatives prior to a more detailed simulation – is discussed in Chapters 3, 4 and later chapters.

As with optimization models, simulation models may be deterministic or stochastic. One of the most useful tools in water resources systems planning is stochastic simulation. While optimization can be used to help define reasonable design and operating policy alternatives to be simulated, simulations can better reveal how each such alternative will perform. Stochastic simulation of complex water resources systems on digital computers provides planners with a way to define the probability distributions of multiple performance indices of those systems.

When simulating any system, the modeller designs an experiment. Initial flow, storage and water quality conditions must be specified if these are being simulated. For example, reservoirs can start full, empty or at random representative conditions. The modeller also determines what data are to be collected on system performance and operation, and how they are to be summarized. The length of time the simulation is to be run must be specified and, in the case of stochastic simulations, the number of runs to be made must also be determined. These considerations are discussed in more detail by Fishman (2001) and in other books on simulation. The use of stochastic simulation and the analysis of the output of such models are introduced here primarily in the context of an example to illustrate what goes into a stochastic simulation model and how one can deal with the information that is generated.

## 9.1. Generating Random Variables

Included in any stochastic simulation model is some provision for the generation of sequences of random numbers that represent particular values of events such as rainfall, streamflows or floods. To generate a sequence of values for a random variable, the probability distribution for the random variable must be specified. Historical data and an understanding of the physical processes are used to select appropriate distributions and to estimate their parameters (as discussed in Section 7.2).

Most computers have algorithms for generating random numbers uniformly distributed (equally likely) between zero and one. This uniform distribution of random numbers is defined by its cdf and pdf;

$$F_U(u) = 0 \quad \text{for} \quad u \le 0,$$

$$= u \quad \text{for} \quad 0 \le u \le 1$$

and

$$= 1 \quad \text{if} \quad u \ge 1 \tag{7.188}$$

so that

$$f_U(u) = 1 \quad \text{if} \quad 0 \le u \le 1 \text{ and } 0 \text{ otherwise} \tag{7.189}$$

These uniform random variables can then be transformed into random variables with any desired distribution. If $F_Q(q_t)$ is the cumulative distribution function of a random variable $Q_t$ in period $t$, then $Q_t$ can be generated using the inverse of the distribution.

$$Q_t = F_Q^{-1}[U_t] \tag{7.190}$$

Here $U_t$ is the uniform random number used to generate $Q_t$. This is illustrated in Figure 7.15.

Analytical expressions for the inverse of many distributions, such as the normal distribution, are not known, so special algorithms are employed to efficiently generate deviates with these distributions (Fishman, 2001).

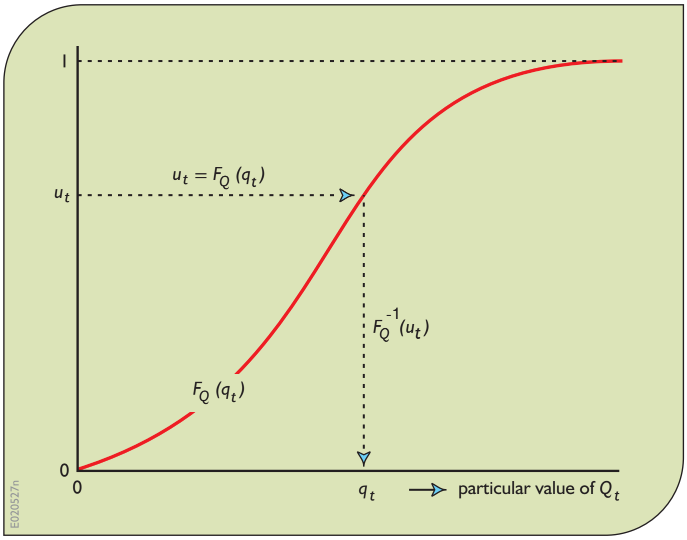



**Figure 7.15.** The probability distribution of a random variable can be inverted to produce values of the random variable.

## 9.2. River Basin Simulation

An example will demonstrate the use of stochastic simulation in the design and analysis of water resources systems. Assume that farmers in a particular valley have been plagued by frequent shortages of irrigation water. They currently draw water from an unregulated river to which they have water rights. A government agency has proposed construction of a moderate-size dam on the river upstream of the points where the farmers withdraw water. The dam would be used to increase the quantity and reliability of irrigation water available to the farmers during the summer growing season.

After preliminary planning, a reservoir with an active capacity of $4 \times 10^7\,m^3$ has been proposed for a natural dam site. It is anticipated that, because of the increased reliability and availability of irrigation water, the quantity of water desired will grow from an initial level of $3 \times 10^7\,m^3/yr$ after construction of the dam to $4 \times 10^7\,m^3/yr$ within six years. After that, demand will grow more slowly to $4.5 \times 10^7\,m^3/yr$ – the estimated maximum reliable yield. The projected demand for summer irrigation water is shown in Table 7.12.

A simulation study can evaluate how the system will be expected to perform over a twenty-year planning period. Table 7.13 contains statistics that describe the hydrology at the dam site. The estimated moments are computed from the forty-five-year historic record.

| year | water demand ($\times 10^7\,m^3/yr$) |
|---|---|
| 1 | 3.0 |
| 2 | 3.2 |
| 3 | 3.4 |
| 4 | 3.6 |
| 5 | 3.8 |
| 6 | 4.0 |
| 7 | 4.1 |
| 8 | 4.2 |
| 9 | 4.3 |
| 10 | 4.3 |
| 11 | 4.4 |
| 12 | 4.4 |
| 13 | 4.4 |
| 14 | 4.4 |
| 15 | 4.5 |
| 16 | 4.5 |
| 17 | 4.5 |
| 18 | 4.5 |
| 19 | 4.5 |
| 20 | 4.5 |

E021101n

**Table 7.12.** Projected water demand for irrigation water.

Using the techniques discussed in the previous section, a Thomas–Fiering model is used to generate twenty-five lognormally distributed synthetic streamflow sequences. The statistical characteristics of the synthetic flows are those listed in Table 7.14. Use of only the forty-five-year historic flow sequence would not allow examination of the system's performance over the large range of streamflow sequences, which could occur during the twenty-year planning period. Jointly, the synthetic sequences should be a description of the range of inflows that the system might experience. A larger number of sequences could be generated.

**Table 7.13.** Characteristics of the river flow.

| | winter | summer | annual | |
|---|---|---|---|---|
| **mean flow** | 4.0 | 2.5 | 6.5 | x $10^7 m^3$ |
| **standard deviation** | 1.5 | 1.0 | 2.3 | x $10^7 m^3$ |
| **correlation of flows:** | | | | |
| winter with following summer | | 0.65 | | |
| summer with following winter | | 0.60 | | |

E021101p

## 9.3. The Simulation Model

The simulation model is composed primarily of continuity constraints and the proposed operating policy. The volume of water stored in the reservoir at the beginning of seasons 1 (winter) and 2 (summer) in year $y$ are denoted by $S_{1y}$ and $S_{2y}$. The reservoir's winter operating policy is to store as much of the winter's inflow $Q_{1y}$ as possible. The winter release $R_{1y}$ is determined by the rule

$$R_{1y} = \begin{cases} S_{1y} + Q_{1y} - K & \text{if} \quad S_{1y} + Q_{1y} - R_{\min} > K \\ R_{\min} & \text{if} \quad K \geq S_{1y} + Q_{1y} - R_{\min} \geq 0 \\ S_{1y} + Q_{1y} & \text{otherwise} \end{cases} \tag{7.191}$$

where $K$ is the reservoir capacity of $4 \times 10^7 \text{ m}^3$ and $R_{\min}$ is $0.50 \times 10^7 \text{ m}^3$, the minimum release to be made if possible. The volume of water in storage at the beginning of the year's summer season is

$$S_{2y} = S_{1y} + Q_{1y} - R_{1y} \tag{7.192}$$

The summer release policy is to meet each year's projected demand or target release $D_y$, if possible, so that

$$\begin{aligned} R_{2y} &= S_{2y} + Q_{2y} - K && \text{if} \quad S_{2y} + Q_{2y} - D_y > K \\ &= D_y && \text{if} \quad 0 \leq S_{2y} + Q_{2y} - D_y \leq K \\ &= S_{2y} + Q_{2y} && \text{otherwise} \end{aligned} \tag{7.193}$$

This operating policy is illustrated in Figure 7.16.

The volume of water in storage at the beginning of the next winter season is

$$S_{1,y+1} = S_{2y} + Q_{2y} - R_{2y} \tag{7.194}$$

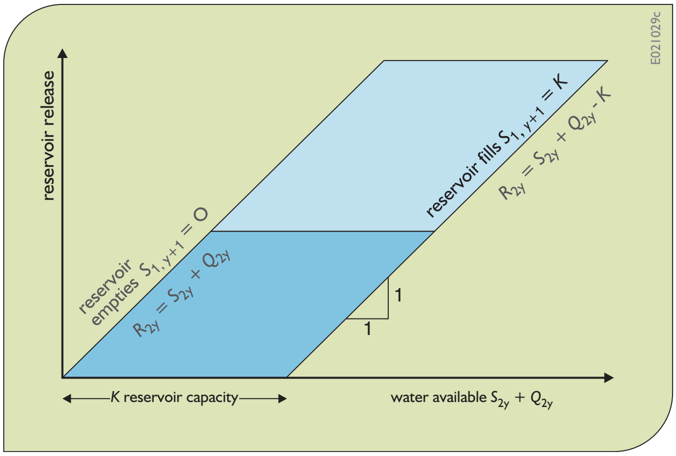



**Figure 7.16.** Summer reservoir operating policy. The shaded area denotes the feasible region of reservoir releases.

## 9.4. Simulation of the Basin

The question to be addressed by this simulation study is how well the reservoir will meet the farmers' water requirements. Three steps are involved in answering this question. First, one must define the performance criteria or indices to be used to describe the system's performance. The appropriate indices will, of course, depend on the problem at hand and the specific concerns of the users and managers of a water resources system. In this example of a reservoir-irrigation system, several indices will be used relating to the reliability with which target releases are met and the severity of any shortages.

The second step is to simulate the proposed system to evaluate the specified indices. For our reservoir-irrigation system, the reservoir's operation was simulated twenty-five times using the twenty-five synthetic streamflow sequences, each twenty years in length. Each of the twenty simulated years consisted of first a winter and then a summer season. At the beginning of the first winter season, the reservoir was taken to be empty ($S_{1y} = 0$ for $y = 1$) because construction would just have been completed. The target release or demand for water in each year is given in Table 7.13.

The third and final step, after simulating the system, is to interpret the resulting information so as to gain an understanding of how the system might perform both with the proposed design and operating policy and with modifications in either the system's design or its operating policy. To see how this may be done, consider the operation of our example reservoir-irrigation system.

The reliability $p_y$ of the target release in year $y$ is the probability that the target release $D_y$ is met or exceeded in that year:

$$p_y = \Pr[R_{2y} \geq D_y] \tag{7.195}$$

The system's reliability is a function of the target release $D_y$, the hydrology of the river, the size of the reservoir and the operating policy of the system. In this example, the reliability also depends on the year in question. Figure 7.17 shows the total number of failures that occurred in each year of the twenty-five simulations. In three of these, the reservoir did not contain sufficient water after the initial winter season to meet the demand the first summer. After year 1, few failures occur in years 2 through 9 because of the low demand. Surprisingly few failures occur in years 10 and 13, when demand has reached its peak; this is because the reservoir was normally full at the beginning of this period as a result of lower demand in the earlier years. Starting in years 14 and after, failures occurred more frequently because of the higher demand placed on the system. Thus one has a sense of how the reliability of the target releases changes over time.

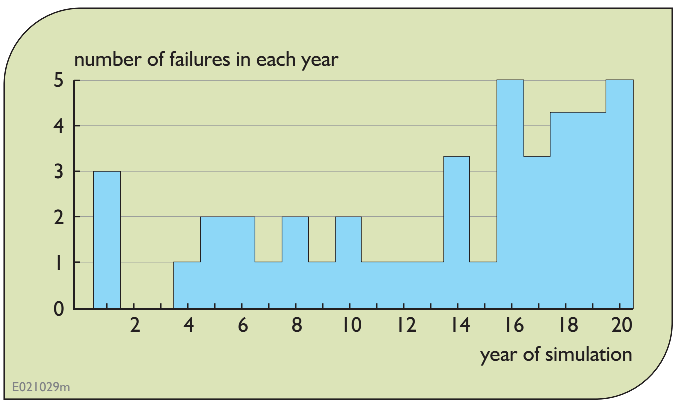



**Figure 7.17.** Number of failures in each year of twenty-five twenty-year simulations.

## 9.5. Interpreting Simulation Output

Table 7.14 contains several summary statistics of the twenty-five simulations. Column 2 of the table contains the average failure frequency in each simulation, which equals the number of years the target release was not met divided by twenty, the number of years simulated. At the bottom of column 2 and the other columns are several statistics that summarize the twenty-five values of the different performance indices. The sample estimates of the mean and variance of each index are given as one way of summarizing the distribution of the observations. Another approach is specification of the sample median, the approximate inter-quartile range $x_{(6)}-x_{(20)}$, and/or the range $x_{(1)}-x_{(25)}$ of the observations, where $x_{(i)}$ is the $i$th largest observation. Either set of statistics could be used to describe the centre and spread of each index's distribution.

Suppose that one is interested in the distribution of the system's failure frequency or, equivalently, the reliability with which the target can be met. Table 7.14 reports that the mean failure rate for the twenty-five simulations is 0.084, implying that the average reliability over the twenty-year period is $1 - 0.084 = 0.916$, or 92%. The median failure rate is 0.05, implying a median reliability of 95%. These are both reasonable estimates of the centre of the distribution of the failure frequency. Note that the actual failure frequency ranged from 0 (seven times) to 0.30. Thus the system's reliability ranged from 100% to as low as 70%, 75% and 80% in runs 17, 8, and 11, respectively. Obviously, the farmers are interested not only in knowing the mean failure frequency but also the range of failure frequencies they are likely to experience.

If one knew the form of the distribution function of the failure frequency, one could use the mean and standard deviation of the observations to determine an interval within which the observations would fall with some pre-specified probability. For example, if the observations are normally distributed, there is a 90% probability that the index falls within the interval $\mu_x \pm 1.65\sigma_x$. Thus, if the simulated failure rates are normally distributed, then there is about a 90% probability

**Table 7.14.** Results of 25 twenty-year simulations.

E021101q

| simulation number, *i* | frequency of failure to meet: target | 80% of target | total shortage TS $x10^7 m^3$ | average deficit, AD |
|---|---|---|---|---|
| 1 | 0.10 | 0.0 | 1.25 | 0.14 |
| 2 | 0.15 | 0.05 | 1.97 | 0.17 |
| 3 | 0.10 | 0.05 | 1.79 | 0.20 |
| 4 | 0.10 | 0.05 | 1.67 | 0.22 |
| 5 | 0.05 | 0.0 | 0.21 | 0.05 |
| 6 | 0.0 | 0.0 | 0.00 | 0.00 |
| 7 | 0.15 | 0.05 | 1.29 | 0.10 |
| 8 | 0.25 | 0.10 | 4.75 | 0.21 |
| 9 | 0.0 | 0.0 | 0.00 | 0.00 |
| 10 | 0.10 | 0.0 | 0.34 | 0.04 |
| 11 | 0.20 | 0.0 | 1.80 | 0.11 |
| 12 | 0.05 | 0.05 | 1.28 | 0.43 |
| 13 | 0.05 | 0.0 | 0.53 | 0.12 |
| 14 | 0.10 | 0.0 | 0.88 | 0.11 |
| 15 | 0.15 | 0.05 | 1.99 | 0.15 |
| 16 | 0.05 | 0.0 | 0.23 | 0.05 |
| 17 | 0.30 | 0.05 | 2.68 | 0.10 |
| 18 | 0.10 | 0.0 | 0.76 | 0.08 |
| 19 | 0.0 | 0.0 | 0.00 | 0.00 |
| 20 | 0.0 | 0.0 | 0.00 | 0.00 |
| 21 | 0.0 | 0.0 | 0.00 | 0.00 |
| 22 | 0.05 | 0.05 | 1.47 | 0.33 |
| 23 | 0.0 | 0.0 | 0.00 | 0.00 |
| 24 | 0.0 | 0.0 | 0.00 | 0.00 |
| 25 | 0.05 | 0.0 | 0.19 | 0.04 |
| mean $\bar{x}$ | 0.084 | 0.020 | 1.00 | 0.106 |
| standard deviation of values; $s_x$ | 0.081 | 0.029 | 1.13 | 0.110 |
| median | 0.05 | 0.00 | 0.76 | 0.10 |
| approximate interquartile range; $x_{(6)}$ - $x_{(20)}$ | 0.0 - 0.15 | 0.0 - 0.05 | 0.0 - 1.79 | 0.0 - 0.17 |
| range; $x_{(1)}$ - $x_{(25)}$ | 0.0 - 0.30 | 0.0 - 0.10 | 0.0 - 4.75 | 0.0 - 0.43 |

that the actual failure rate observed in any simulation is within the interval $\bar{x} \pm 1.65 s_x$. In our case this interval would be [0.084 − 1.65(0.081), 0.084 + 1.65(0.081)] = [−0.050, 0.218]. Clearly, the failure rate cannot be less than zero, so this interval makes little sense in our example.

A more reasonable approach to describing the distribution of a performance index whose probability distribution function is not known is to use the observations themselves. If the observations are of a continuous random variable, the interval $x_{(i)} - x_{(n+1-i)}$ provides a reasonable estimate of an interval within which the random variable falls with probability

$$P = \frac{n+1-i}{n+1} - \frac{i}{n+1} = \frac{n+1-2i}{n+1} \tag{7.196}$$

In our example, the range $x_{(1)} - x_{(25)}$ of the twenty-five observations is an estimate of an interval in which a continuous random variable falls with probability (25 + 1 − 2)/(25 + 1) = 92%, while $x_{(6)} - x_{(20)}$ corresponds to probability (25 + 1 − 2 × 6)/(25 + 1) = 54%.

Table 7.14 reports that, for the failure frequency, $x_{(1)} - x_{(25)}$ equals 0−0.30, while $x_{(6)} - x_{(20)}$ equals 0−0.15. Reflection on how the failure frequencies are calculated reminds us that the failure frequency can only take on the discrete, non-negative values 0, 1/20, 2/20, … , 20/20. Thus, the random variable $X$ cannot be less than zero. Hence, if the lower endpoint of an interval is zero, as is the case here, then $0 - x_{(k)}$ is an estimate of an interval within which the random variable falls with a probability of at least $k/(n + 1)$. For $k$ equal to 20 and 25, the corresponding probabilities are 77% and 96%.

Often, the analysis of a simulated system's performance centres on the average value of performance indices, such as the failure rate. It is important to know the accuracy with which the mean value of an index approximates the true mean. This is done by the construction of *confidence intervals*. A confidence interval is an interval that will contain the unknown value of a parameter with a specified probability. Confidence intervals for a mean are constructed using the $t$ statistic,

$$t = \frac{\bar{x} - \mu_x}{s_x/\sqrt{n}} \tag{7.197}$$

which, for large $n$, has approximately a standard normal distribution. Certainly, $n = 25$ is not very large, but the approximation to a normal distribution may be sufficiently good to obtain a rough estimate of how close the average frequency of failure $\bar{x}$ is likely to be to $\mu_x$. A 100(1 − 2$\alpha$)% confidence interval for $\mu_x$ is, approximately,

$$\bar{x} - t_\alpha \frac{s_x}{\sqrt{n}} \le \mu_x \le \bar{x} + t_\alpha \frac{s_x}{\sqrt{n}}$$

or

$$0.084 - t_\alpha\left(\frac{0.081}{\sqrt{25}}\right) \le \mu_x \le 0.084 + t_\alpha\left(\frac{0.081}{\sqrt{25}}\right) \tag{7.198}$$

If $\alpha = 0.05$, then using a normal distribution $t_\alpha = 1.65$ and Equation 7.118 becomes $0.057 \le \mu_x \le 0.11$.

Hence, based on the simulation output, one can be about 90% sure that the true mean failure frequency lies between 5.7% and 11%. This corresponds to a reliability of between 89% and 94%. By performing additional simulations to increase the size of $n$, the width of this confidence interval can be decreased. However, this increase in accuracy may be an illusion because the uncertainty in the parameters of the streamflow model has not been incorporated into the analysis.

Failure frequency or system reliability describes only one dimension of the system's performance. Table 7.14 contains additional information on the system's performance related to the severity of shortages. Column 3 lists the frequencies with which the shortage exceeded 20% of that year's demand. This occurred in approximately 2% of the years, or in 24% of the years in which a failure occurred. Taking another point of view, failures in excess of 20% of demand occurred in nine out of twenty-five, or in 36% of the simulation runs.

Columns 4 and 5 of Table 7.14 contain two other indices that pertain to the severity of the failures. The total shortfall, $TS$, in Column 4 is calculated as the sum of the positive differences between the demand and the release in the summer season over the twenty-year period.

$$TS = \sum_y [D_{2y} - R_{2y}]^+$$

where

$$[Q]^+ = Q \quad \text{if } Q > 0; \quad 0 \text{ otherwise} \tag{7.199}$$

The total shortfall equals the total amount by which the target release is not met in years in which shortages occur.

Related to the total shortfall is the average deficit. The deficit is defined as the shortfall in any year divided by the target release in that year. The average deficit, AD, is

$$AD = \frac{1}{m}\sum_{y=1}^{20}\frac{\left[D_{2y} - R_{2y}\right]}{D_{2y}} \tag{7.200}$$

where $m$ is the number of failures (deficits) or nonzero terms in the sum.

Both the total shortfall and the average deficit measure the severity of shortages. The mean total shortfall $\overline{TS}$, equal to 1.00 for the twenty-five simulation runs, is a difficult number to interpret. While no shortage occurred in seven runs, the total shortage was 4.7 in run 8, in which the shortfall in two different years exceeded 20% of the target. The median of the total shortage values, equal to 0.76, is an easier number to interpret in that one knows that half the time the total shortage was greater and half the time less than this value.

The mean average deficit $\overline{AD}$ is 0.106, or 11%. However, this mean includes an average deficit of zero in the seven runs in which no shortages occurred. The average deficit in the eighteen runs in which shortages occurred is (11%)(25/18) = 15%. The average deficit in individual simulations in which shortages occurred ranges from 4% to 43%, with a median of 11.5%.

After examining the results reported in Table 7.14, the farmers might determine that the probability of a shortage exceeding 20% of a year's target release is higher than they would like. They can deal with more frequent minor shortages, not exceeding 20% of the target, with little economic hardship, particularly if they are warned at the beginning of the growing season that less than the targeted quantity of water will be delivered. Then they can curtail their planting or plant crops requiring less water.

In an attempt to find out how better to meet the farmers' needs, the simulation program was re-run with the same streamflow sequences and a new operating policy in which only 80% of the growing season's target release is provided (if possible) if the reservoir is less than 80% full at the end of the previous winter season. This gives the farmers time to adjust their planting schedules and may increase the quantity of water stored in the reservoir to be used the following year if the drought persists.

As the simulation results with the new policy in Table 7.15 demonstrate, this new operating policy appears to have the expected effect on the system's operation. With the new policy, only six severe shortages in excess of 20% of demand occur in the twenty-five twenty-year simulations, as opposed to ten such shortages with the original policy. In addition, these severe shortages are all less severe than the corresponding shortages that occur with the same streamflow sequence when the original policy is followed.

The decrease in the severity of shortages is obtained at a price. The overall failure frequency has increased from 8.4% to 14.2%. However, the latter value is misleading because in fourteen of the twenty-five simulations, a failure occurs in the first simulation year with the new policy, whereas only three failures occur with the original policy. Of course, these first-year failures occur because the reservoir starts empty at the beginning of the first winter and often does not fill that season. Ignoring these first-year failures, the failure rates with the two policies over the subsequent nineteen years are 8.2% and 12.0%. Thus the frequency of failures in excess of 20% of demand is decreased from 2.0% to 1.2% by increasing the frequency of all failures after the first year from 8.2% to 12.0%. Reliability decreases, but so does vulnerability. If the farmers are willing to put up with more frequent minor shortages, then it appears that they can reduce their risk of experiencing shortages of greater severity.

The preceding discussion has ignored the statistical issue of whether the differences between the indices obtained in the two simulation experiments are of sufficient statistical reliability to support the analysis. If care is not taken, observed changes in a performance index from one simulation experiment to another may be due to sampling fluctuations rather than to modifications of the water resource system's design or operating policy.

As an example, consider the change that occurred in the frequency of shortages. Let $X_{1i}$ and $X_{2i}$ be the simulated failure rates using the $i$th streamflow sequence with the original and modified operating policies. The random variables $Y_i = X_{1i} - X_{2i}$ for $i$ equal 1 through 25 are independent of each other if the streamflow sequences are generated independently, as they were.

One would like to confirm that the random variable $Y$ tends to be negative more often than it is positive, and hence, that policy 2 indeed results in more failures overall. A direct test of this theory is provided by the sign test. Of the twenty-five paired simulation runs, $y_i < 0$ in twenty-one cases and $y_i = 0$ in four cases. We can ignore the times

E021101r

| simulation number, $i$ | frequency of failure to meet: target | 80% of target | total shortage TS $x10^7 m^3$ | average deficit, AD |
|---|---|---|---|---|
| 1 | 0.10 | 0.0 | 1.80 | 0.20 |
| 2 | 0.30 | 0.0 | 4.70 | 0.20 |
| 3 | 0.25 | 0.0 | 3.90 | 0.20 |
| 4 | 0.20 | 0.05 | 3.46 | 0.21 |
| 5 | 0.10 | 0.0 | 1.48 | 0.20 |
| 6 | 0.05 | 0.0 | 0.60 | 0.20 |
| 7 | 0.20 | 0.0 | 3.30 | 0.20 |
| 8 | 0.25 | 0.10 | 5.45 | 0.26 |
| 9 | 0.05 | 0.0 | 0.60 | 0.20 |
| 10 | 0.20 | 0.0 | 3.24 | 0.20 |
| 11 | 0.25 | 0.0 | 3.88 | 0.20 |
| 12 | 0.10 | 0.05 | 1.92 | 0.31 |
| 13 | 0.10 | 0.0 | 1.50 | 0.20 |
| 14 | 0.15 | 0.0 | 2.52 | 0.20 |
| 15 | 0.25 | 0.05 | 3.76 | 0.18 |
| 16 | 0.10 | 0.0 | 1.80 | 0.20 |
| 17 | 0.30 | 0.0 | 5.10 | 0.20 |
| 18 | 0.15 | 0.0 | 2.40 | 0.20 |
| 19 | 0.0 | 0.0 | 0.0 | 0.0 |
| 20 | 0.05 | 0.0 | 0.76 | 0.20 |
| 21 | 0.10 | 0.0 | 1.80 | 0.20 |
| 22 | 0.10 | 0.05 | 2.37 | 0.26 |
| 23 | 0.05 | 0.0 | 0.90 | 0.20 |
| 24 | 0.05 | 0.0 | 0.90 | 0.20 |
| 25 | 0.10 | 0.0 | 1.50 | 0.20 |
| mean $\bar{x}$ | 0.142 | 0.012 | 2.39 | 0.201 |
| standard deviation of values; $s_x$ | 0.087 | 0.026 | 1.50 | 0.050 |
| median | 0.10 | 0.00 | 1.92 | 0.20 |
| approximate interquartile range; $x_{(6)} - x_{(20)}$ | 0.05 - 0.25 | 0.0 - 0.0 | 0.90 - 3.76 | 0.20 - 0.20 |
| range; $x_{(1)} - x_{(25)}$ | 0.0 - 0.30 | 0.0 - 0.10 | 0.0 - 5.45 | 0.0 - 0.31 |

**Table 7.15.** Results of 25 twenty-year simulations with modified operating policy to avoid severe shortages.

when $y_i = 0$. Note that if $y_i < 0$ and $y_i > 0$ were equally likely, then the probability of observing $y_i < 0$ in all twenty-one cases when $y_i \neq 0$ is $2^{-21}$ or $5 \times 10^{-7}$. This is exceptionally strong proof that the new policy has increased the failure frequency.

A similar analysis can be made of the frequency with which the release is less than 80% of the target. Failure frequencies differ in the two policies in only four of the twenty-five simulation runs. However, in all four cases where they differ, the new policy resulted in fewer severe failures. The probability of such a lopsided result, were it equally likely that either policy would result in a lower frequency of failures in excess of 20% of the target, is $2^{-4} = 0.0625$. This is fairly strong evidence that the new policy indeed decreases the frequency of severe failures.

Another approach to this problem is to ask if the difference between the average failure rates $\bar{x}_1$ and $\bar{x}_2$ is statistically significant; that is, can the difference between $X_1$ and $X_2$ be attributed to the fluctuations that occur in the average of any finite set of random variables? In this example the significance of the difference between the two means can be tested using the random variable $Y_i$ defined as $X_{1i} - X_{2i}$ for $i$ equal 1 through 25. The mean of the observed $y_i$'s is

$$\begin{aligned}\bar{y} &= \frac{1}{25}\sum_{i-1}^{25}(x_{1i} - x_{2i}) = \bar{x}_1 - \bar{x}_2 \\ &= 0.084 - 0.142 = -0.058\end{aligned} \tag{7.201}$$

and their variance is

$$s_y^2 = \frac{1}{25}\sum_{i=1}^{25}(x_{1i} - x_{2i} - \bar{y})^2 = (0.0400)^2 \tag{7.202}$$

Now, if the sample size $n$, equal to 25 here, is sufficiently large, then $t$ defined by

$$t = \frac{\bar{y} - \mu_Y}{s_Y/\sqrt{n}} \tag{7.203}$$

has approximately a standard normal distribution. The closer the distribution of $Y$ is to that of the normal distribution, the faster the convergence of the distribution of $t$ is to the standard normal distribution with increasing $n$. If $X_{1i} - X_{2i}$ is normally distributed, which is not the case here, then each $Y_i$ has a normal distribution and $t$ has Student's $t$-distribution.

If $E[X_{1i}] = E[X_{2i}]$, then $\mu_Y$ equals zero, and upon substituting the observed values of $\bar{y}$ and $s_Y^2$ into Equation 7.123, one obtains

$$t = \frac{-0.0580}{0.0400/\sqrt{25}} = -7.25 \tag{7.204}$$

The probability of observing a value of $t$ equal to $-7.25$ or smaller is less than 0.1% if $n$ is sufficiently large that $t$ is normally distributed. Hence it appears very improbable that $\mu_Y$ equals zero.

This example provides an illustration of the advantage of using the same streamflow sequences when simulating both policies. Suppose that different streamflow sequences were used in all the simulations. Then the expected value of $Y$ would not change, but its variance would be given by

$$\begin{aligned}\mathrm{Var}(Y) &= E[X_1 - X_2 - (\mu_1 - \mu_2)]^2 \\ &= E[(X_1 - \mu_1)^2] - 2E[(X_1 - \mu_1) \\ &\quad \times (X_2 - \mu_2)] + E[(X_2 - \mu_2)^2] \\ &= \sigma_{x_1}^2 - 2\mathrm{Cov}(X_1, X_2) + \sigma_{x_2}^2\end{aligned} \tag{7.205}$$

where $\mathrm{Cov}(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)]$ and is the *covariance* of the two random variables. The covariance between $X_1$ and $X_2$ will be zero if they are independently distributed, as they would be if different randomly generated streamflow sequences were used in each simulation. Estimating $\sigma_{x_1}^2$ and $\sigma_{x_2}^2$ by their sample estimates, an estimate of what the variance of $Y$ would be if $\mathrm{Cov}(X_1, X_2)$ were zero is

$$\hat{\sigma}_Y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 = (0.081)^2 + (0.087)^2 = (0.119)^2 \tag{7.206}$$

The actual sample estimate $\sigma_Y$ equals 0.040; if independent streamflow sequences are used in all simulations, $\sigma_Y$ will take a value near 0.119 rather than 0.040 (Equation 7.202). A standard deviation of 0.119, with $\mu_y = 0$, yields a value of the $t$ test statistic

$$t = \frac{\bar{y} - \mu_Y}{0.119/\sqrt{25}} = -2.44 \tag{7.207}$$

If $t$ is normally distributed, the probability of observing a value less than $-2.44$ is about 0.8%. This illustrates that use of the same streamflow sequences in the simulation of both policies allows one to better distinguish the differences in the policies' performance. By using the same streamflow sequences, or other random inputs, one can construct a simulation experiment in which variations in performance caused by different random inputs are confused as little as possible with the differences in performance caused by changes in the system's design or operating policy.

# 10. Conclusions

This chapter has introduced statistical concepts that analysts use to describe the randomness or uncertainty of their data. Most of the data used by water resources systems analysts is uncertain. This uncertainty comes from not understanding as well as we would like how our water resources systems (including their ecosystems) function as well as not being able to forecast, perfectly, the future. It is that simple. We do not know the exact amounts, qualities and their distributions over space and time of either the supplies of water we manage or the water demands we try to meet. We also do not know the benefits and costs, however measured, of any actions we take to manage both water supply and water demand.

The chapter began with an introduction to probability concepts and methods for describing random variables and parameters of their distributions. It then reviewed some of the commonly used probability distributions and how to determine the distributions of sample data, how to work with censored and partial duration series data, methods of regionalization, stochastic processes and time-series analyses.

The chapter concluded with an introduction to a range of univariate and multivariate stochastic models that are used to generate stochastic streamflow, precipitation depths, temperatures and evaporation. These methods are used to generate temporal and spatial stochastic process that serve as inputs to stochastic simulation models for system design, for system operations studies, and for evaluation of the reliability and precision of different estimation algorithms. The final section of this chapter provides an example of stochastic simulation, and the use of statistical methods to summarize the results of such simulations.

This is merely an introduction to some of the statistical tools available for use when dealing with uncertain data. Many of the concepts introduced in this chapter will be used in the chapters that follow on constructing and implementing various types of optimization, simulation and statistical models. The references cited in the reference section provide additional and more detailed information.

Although many of the methods presented in this and in some of the following chapters can describe many of the characteristics and consequences of uncertainty, it is unclear as to whether or not society knows exactly what to do with such information. Nevertheless, there seems to be an increasing demand from stakeholders involved in planning processes for information related to the uncertainty associated with the impacts predicted by models. The challenge is not only to quantify that uncertainty, but also to communicate it in effective ways that inform, and not confuse, the decision-making process.

# 11. References


AYYUB, B.M. and MCCUEN, R.H. 2002. *Probability, statistics, and reliability for engineers and scientists*. Boca Raton, Chapman and Hill, CRC Press.

BARTOLINI, P. and SALAS, J. 1993. Modelling of streamflow processes at different time scales. *Water Resources Research*, Vol. 29, No. 8, pp. 2573–87.

BATES, B.C. and CAMPBELL, E.P. 2001. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall–runoff modelling. *Water Resources Research*, Vol. 37, No. 4, pp. 937–47.

BEARD, L.R. 1960. Probability estimates based on small normal-distribution samples. *Journal of Geophysical Research*, Vol. 65, No. 7, pp. 2143–8.

BEARD, L.R. 1997. Estimating flood frequency and average annual damages. *Journal of Water Resources Planning and Management*, Vol. 123, No. 2, pp. 84–8.

BENJAMIN, J.R. and CORNELL, C.A. 1970. *Probability, statistics and decisions for civil engineers*. New York, McGraw-Hill.

BICKEL, P.J. and DOKSUM, K.A. 1977. *Mathematical statistics: basic ideas and selected topics*. San Francisco, Holden-Day.

BOBÉE, B. 1975. The log Pearson type 3 distribution and its applications in hydrology. *Water Resources Research*, Vol. 14, No. 2, pp. 365–9.

BOBÉE, B. and ASHKAR, F. 1991. *The gamma distribution and derived distributions applied in hydrology*. Littleton Colo., Water Resources Press.

BOBÉE, B. and ROBITAILLE, R. 1977. The use of the Pearson type 3 distribution and log Pearson type 3 distribution revisited. *Water Resources Research*, Vol. 13, No. 2, pp. 427–43.

BOX, G.E.P.; JENKINS, G.M. and RISINSEL, G.C. 1994. *Times series analysis: forecasting and control*, 3rd Edition. New Jersey, Prentice-Hall.

CAMACHO, F.; MCLEOD, A.I. and HIPEL, K.W. 1985. Contemporaneous autoregressive-moving average (CARMA) modelling in water resources. *Water Resources Bulletin*, Vol. 21, No. 4, pp. 709–20.

CARLIN, B.P. and LOUIS, T.A. 2000. *Bayes and empirical Bayes methods for data analysis*, 2nd Edition. New York, Chapman and Hall, CRC.

CHOWDHURY, J.U. and STEDINGER, J.R. 1991. Confidence intervals for design floods with estimated skew coefficient. *Journal of Hydraulic Engineering*, Vol. 117, No. 7, pp. 811–31.

CHOWDHURY, J.U.; STEDINGER, J.R. and LU, L.H. 1991. Goodness-of-fit tests for regional GEV flood distributions. *Water Resources Research*, Vol. 27, No. 7, pp. 1765–76.

CLAPS, P. 1993. Conceptual basis for stochastic models of monthly streamflows. In: J.B. Marco, R. Harboe and J.D. Salas (eds). *Stochastic hydrology and its use in water resources systems simulation and optimization*, Dordrecht, Kluwer Academic, pp. 229–35.

CLAPS, P.; ROSSI, F. and VITALE, C. 1993. Conceptual-stochastic modelling of seasonal runoff using autoregressive moving average models and different scales of aggregation. *Water Resources Research*, Vol. 29, No. 8, pp. 2545–59.

COHN, T.A.; LANE, W.L. and BAIER, W.G. 1997. An algorithm for computing moments-based flood quantile estimates when historical flood information is available. *Water Resources Res*earch, Vol. 33, No. 9, pp. 2089–96.

COHN, C.A.; LANE, W.L. and STEDINGER, J.R. 2001. Confidence intervals for EMA flood quantile estimates. *Water Resources Research*, Vol. 37, No. 6, pp. 1695–1706.

CRAINICEANU, C.M.; RUPPERT, D.; STEDINGER, J.R. and BEHR, C.T. 2002. *Improving MCMC mixing for a GLMM describing pathogen concentrations in water supplies, in case studies in Bayesian analysis*. New York, Springer-Verlag.

D'AGOSTINO, R.B. and STEPHENS, M.A. 1986. *Goodness-of-fit procedures*. New York, Marcel Dekker.

DAVID, H.A. 1981. *Order statistics,* 2nd edition. New York, Wiley.

FIERING, M.B. 1967. *Streamflow synthesis*. Cambridge, Mass., Harvard University Press.

FILL, H. and STEDINGER, J. 1995. L-moment and PPCC goodness-of-fit tests for the Gumbel distribution and effect of autocorrelation. *Water Resources Research,* Vol. 31, No. 1, pp. 225–29.

FILL, H. and STEDINGER, J. 1998. Using regional regression within index flood procedures and an empirical Bayesian estimator. *Journal of Hydrology*, Vol. 210, Nos 1–4, pp. 128–45.

FILLIBEN, J.J. 1975. The probability plot correlation test for normality. *Technometrics*, Vol. 17, No. 1, pp. 111–17.

FISHMAN, G.S. 2001. *Discrete-event simulation: modelling, programming, and analysis*. Berlin, Springer-Verlag.

GABRIELE, S. and ARNELL, N. 1991. A hierarchical approach to regional flood frequency analysis. *Water Resources Research*, Vol. 27, No. 6, pp. 1281–9.

GELMAN, A.; CARLIN, J.B.; STERN, H.S. and RUBIN, D.B. 1995. *Bayesian data analysis*. Boca Raton, Chapman and Hall, CRC.

GILKS, W.R.; RICHARDSON, S. and SPIEGELHALTER, D.J. (eds). 1996. *Markov chain Monte Carlo in practice*. London and New York, Chapman and Hall.

GIESBRECHT, F. and KEMPTHORNE, O. 1976. Maximum likelihood estimation in the three-parameter log normal distribution. *Journal of the Royal Statistical Society B*, Vol. 38, No. 3, pp. 257–64.

GREENWOOD, J.A. and DURAND, D. 1960. Aids for fitting the gamma distribution by maximum likelihood. *Technometrics*, Vol. 2, No. 1, pp. 55–65.

GRIFFS, V.W.; STEDINGER, J.R. and COHN, T.A. 2004. LP3 quantile estimators with regional skew information and low outlier adjustments. *Water Resources Research*, Vol. 40, forthcoming.

GRYGIER, J.C. and STEDINGER, J.R. 1988. Condensed disaggregation procedures and conservation corrections. *Water Resources Research*, Vol. 24, No. 10, pp. 1574–84.

GRYGIER, J.C. and STEDINGER, J.R. 1991. *SPIGOT: a synthetic flow generation software package, user's manual and technical description, version 2.6*. Ithaca, N.Y., School of Civil and Environmental Engineering, Cornell University.

GUMBEL, E.J. 1958. *Statistics of extremes*. New York, Columbia University Press.

GUPTA, V.K. and DAWDY, D.R. 1995a. Physical interpretation of regional variations in the scaling exponents of flood quantiles. *Hydrological Processes*, Vol. 9, Nos. 3–4, pp. 347–61.

GUPTA, V.K. and DAWDY, D.R. 1995b. Multiscaling and skew separation in regional floods. *Water Resources Research*, Vol. 31, No. 11, pp. 2761–76.

GUPTA, V.K.; MESA, O.J. and DAWDY, D.R. 1994. Multiscaling theory of flood peaks: regional quantile analysis. *Water Resources Research*, Vol. 30, No. 12, pp. 3405–12.

HAAN, C.T. 1977. *Statistical methods in hydrology*. Ames, Iowa, Iowa State University Press.

HAAS, C.N. and SCHEFF, P.A. 1990. Estimation of averages in truncated samples. *Environmental Science and Technol*ogy, Vol. 24, No. 6, pp. 912–19.

HELSEL, D.R. 1990. Less than obvious: statistical treatment of data below the detection limit. *Environ. Sci. and Technol.*, Vol. 24, No. 12, pp. 1767–74.

HELSEL, D.R. and COHN, T.A. 1988. Estimation of descriptive statistics for multiple censored water quality data. *Water Resources Research*, Vol. 24, No. 12, pp. 1997–2004.

HIRSCH, R.M. 1979. Synthetic hydrology and water supply reliability. *Water Resources Research*, Vol. 15, No. 6, pp. 1603–15.

HIRSCH, R.M. and STEDINGER, J.R. 1987. Plotting positions for historical floods and their precision. *Water Resources Research*, Vol. 23, No. 4, pp. 715–27.

HOSHI, K.; BURGES, S.J. and YAMAOKA, I. 1978. Reservoir design capacities for various seasonal operational hydrology models. *Proceedings of the Japanese Society of Civil Engineers*, No. 273, pp. 121–34.

HOSHI, K.; STEDINGER, J.R. and BURGES, S. 1984. Estimation of log normal quantiles: Monte Carlo results and first-order approximations. *Journal of Hydrology*, Vol. 71, Nos 1–2, pp. 1–30.

HOSKING, J.R.M. 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of Royal Statistical Society*, *B,* Vol. 52, No. 2, pp. 105–24.

HOSKING, J.R.M. and WALLIS, J.R. 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, Vol. 29, No. 3, pp. 339–49.

HOSKING, J.R.M. and WALLIS, J.R. 1995. A comparison of unbiased and plotting-position estimators of L-moments. Water Resources Research, Vol. 31, No. 8, pp. 2019–25.

HOSKING, J.R.M. and WALLIS, J.R. 1997. *Regional frequency analysis: an approach based on L-moments*. Cambridge, Cambridge University Press.

HOSKING, J.R.M.; WALLIS, J.R. and WOOD, E.F. 1985. Estimation of the generalized extreme-value distribution by the method of probability weighted moments. *Technometrics*, Vol. 27, No. 3, pp. 251–61.

IACWD (Interagency Advisory Committee on Water Data). 1982. *Guidelines for determining flood flow frequency*, Bulletin 17B. Reston, Va., US Department of the Interior, US Geological Survey, Office of Water Data Coordination.

JENKINS, G.M. and WATTS, D.G. 1968. *Spectral Analysis and its Applications*. San Francisco, Holden-Day.

KENDALL, M.G. and STUART, A. 1966. *The advanced theory of statistics*, Vol. 3. New York, Hafner.

KIRBY, W. 1974. Algebraic boundness of sample statistics. *Water Resources Research*, Vol. 10, No. 2, pp. 220–2.

KIRBY, W. 1972. Computer oriented Wilson–Hilferty transformation that preserves the first 3 moments and lower bound of the Pearson Type 3 distribution. *Water Resources Research*, Vol. 8, No. 5, pp. 1251–4.

KITE, G.W. 1988. *Frequency and risk analysis in hydrology*. Littleton, Colo. Water Resources Publications.

KOTTEGODA, M. and ROSSO, R. 1997. *Statistics, probability, and reliability for civil and environmental engineers*. New York, McGraw-Hill.

KOUTSOYIANNIS, D. and MANETAS, A. 1996. Simple disaggregation by accurate adjusting procedures. *Water Resources Research*, Vol. 32, No. 7, pp. 2105–17.

KROLL, K. and STEDINGER, J.R. 1996. Estimation of moments and quantiles with censored data. *Water Resources Research,* Vol. 32, No. 4, pp. 1005–12.

KUCZERA, G. 1999. Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. Water *Resources Research,* Vol. 35, No. 5, pp. 1551–7.

KUCZERA, G. 1983. Effects of sampling uncertainty and spatial correlation on an empirical Bayes procedure for combining site and regional information. *Journal of Hydrology*, Vol. 65, No. 4, pp. 373–98.

KUCZERA, G. 1982. Combining site-specific and regional information: an empirical Bayesian approach. *Water Resources Research*, Vol. 18, No. 2, pp. 306–14.

LANDWEHR, J.M.; MATALAS, N.C. and WALLIS, J.R. 1978. Some comparisons of flood statistics in real and log space. *Water Resources Research*, Vol. 14, No. 5, pp. 902–20.

LANDWEHR, J.M.; MATALAS, N.C. and WALLIS, J.R. 1979. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, Vol. 15, No. 5, pp. 1055–64.

LANE, W. 1979. *Applied stochastic techniques (users manual)*. Denver, Colo., Bureau of Reclamation, Engineering and Research Center, December.

LANGBEIN, W.B. 1949. Annual floods and the partial duration flood series, EOS. *Transactions of the American Geophysical Union*, Vol. 30, No. 6, pp. 879–81.

LEDOLTER, J. 1978. The analysis of multivariate time series applied to problems in hydrology. *Journal of Hydrology*, Vol. 36, No. 3–4, pp. 327–52.

LETTENMAIER, D.P. and BURGES, S.J. 1977a. Operational assessment of hydrological models of long-term persistence. *Water Resources Research*, Vol. 13, No. 1, pp. 113–24.

LETTENMAIER, D.P. and BURGES, S.J. 1977b. An operational approach to preserving skew in hydrological models of long-term persistence. *Water Resources Research*, Vol. 13, No. 2, pp. 281–90.

LETTENMAIER, D.P.; WALLIS, J.R. and WOOD, E.F. 1987. Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research*, Vol. 23, No. 2, pp. 313–24.

MACBERTHOUEX, P. and BROWN, L.C. 2002. *Statistics for environmental engineers*, 2nd Edition. Boca Raton, Fla., Lewis, CRC Press.

MADSEN, H.; PEARSON, C.P.; RASMUSSEN, P.F. and ROSBJERG, D. 1997a. Comparison of annual maximum series and partial duration series methods for modelling extreme hydrological events 1: at-site modelling. *Water Resources Research*, Vol. 33, No. 4, pp. 747–58.

MADSEN, H.; PEARSON, C.P. and ROSBJERG, D. 1997b. Comparison of annual maximum series and partial duration series methods for modelling extreme hydrological events 2: regional modelling. *Water Resources Research,* Vol. 33, No. 4, pp. 759–70.

MADSEN, H. and ROSBJERG, D. 1997a. The partial duration series method in regional index flood modelling. *Water Resources Research*, Vol. 33, No. 4, pp. 737–46.

MADSEN, H. and ROSBJERG, D. 1997b. Generalized least squares and empirical Bayesian estimation in regional partial duration series index-flood modelling. *Water Resources Research*, Vol. 33, No. 4, pp. 771–82.

MAHEEPALA, S. and PERERA, B.J.C. 1996. Monthly hydrological data generation by disaggregation. *Journal of Hydrology*, No. 178, 277–91.

MARCO, J.B.; HARBOE, R. and SALAS, J.D. (eds). 1989. Stochastic hydrology and its use in water resources systems simulation and optimization. NATO ASI Series. Dordrecht, Kluwer Academic.

MARTINS, E.S. and STEDINGER, J.R. 2000. Generalized maximum likelihood GEV quantile estimators for hydrological data. *Water Resources Research*, Vol. 36, No. 3, pp. 737–44.

MARTINS, E.S. and STEDINGER, J.R. 2001a. Historical information in a GMLE-GEV framework with partial duration and annual maximum series. *Water Resources Research*, Vol. 37, No. 10, pp. 2551–57.

MARTINS, E.S. and STEDINGER, J.R. 2001b. Generalized maximum likelihood Pareto-Poisson flood risk analysis for partial duration series. *Water Resources Research*, Vol. 37, No. 10, pp. 2559–67.

MATALAS, N.C. 1967. Mathematical assessment of synthetic hydrology. *Water Resources Research*, Vol. 3, No. 4, pp. 937–45.

MATALAS, N.C. and WALLIS, J.R. 1973. Eureka! It fits a Pearson type 3 distribution. *Water Resources Research*, Vol. 9, No. 3, pp. 281–9.

MATALAS, N.C. and WALLIS, J.R. 1976. Generation of synthetic flow sequences. In: A.K. Biswas (ed.), *Systems approach to water management*. New York, McGraw-Hill.

MEJIA, J.M. and ROUSSELLE, J. 1976. Disaggregation models in hydrology revisited. *Water Resources Research*, Vol. 12, No. 2, pp. 185–6.

NERC (Natural Environmental Research Council) 1975. Flood studies report, Vol. 1: hydrological studies. London.

NORTH, M. 1980. Time-dependent stochastic model of floods. *Journal of the Hydraulics Division*, ASCE, Vol. 106, No. HY5, pp. 649–65.

O'CONNELL, D.R.H.; OSTENAA, D.A.; LEVISH, D.R. and KLINGER, R.E. 2002. Bayesian flood frequency analysis with paleohydrological bound data. *Water Resources Research*, Vol. 38, No. 5, pp. 161–64.

O'CONNELL, P.E. 1977. ARMA models in synthetic hydrology. In: T.A. Ciriani, V. Maione and J.R. Wallis (eds), *Mathematical models for surface water hydrology*. New York, Wiley.

PRESS, W.H.; FLANNERY, B.P.; TEUKOLSKY, S.A. and VETTERLING, W.T. 1986. *Numerical recipes: the art of scientific computing*. Cambridge, UK, Cambridge University Press.

RAIFFA, H. and SCHLAIFER, R. 1961. *Applied statistical decision theory*. Cambridge, Mass., MIT Press.

RASMUSSEN, P.F. and ROSBJERG, D. 1989. Risk estimation in partial duration series. *Water Resources Research*, Vol. 25, No. 11, pp. 2319–30.

RASMUSSEN, P.F. and ROSBJERG, D. 1991a. Evaluation of risk concepts in partial duration series. *Stochastic Hydrology and Hydraulics*, Vol. 5, No. 1, pp. 1–16.

RASMUSSEN, P.F. and ROSBJERG, D. 1991b. Prediction uncertainty in seasonal partial duration series. *Water Resources Research*, Vol. 27, No. 11, pp. 2875–83.

RASMUSSEN, R.F.; SALAS, J.D.; FAGHERAZZI, L.; RASSAM.; J.C. and BOBÉE, R. 1996. Estimation and validation of contemporaneous PARMA models for streamflow simulation. *Water Resources Research*, Vol. 32, No. 10, pp. 3151–60.

ROBINSON, J.S. and SIVAPALAN, M. 1997. Temporal scales and hydrological regimes: implications for flood frequency scaling. *Water Resources Research*, Vol. 33, No. 12, pp. 2981–99.

ROSBJERG, D. 1985. Estimation in Partial duration series with independent and dependent peak values. *Journal of Hydrology*, *No.* 76, pp. 183–95.

ROSBJERG, D. and MADSEN, H. 1998. Design with uncertain design values, In: H. Wheater and C. Kirby (eds), *Hydrology in a changing environment*, New York, Wiley. Vol. 3, pp. 155–63.

ROSBJERG, D.; MADSEN, H. and RASMUSSEN, P.F. 1992. Prediction in partial duration series with generalized Pareto-distributed exceedances. *Water Resources Research*, Vol. 28, No. 11, pp. 3001–10.

SALAS, J.D. 1993. Analysis and modelling of hydrological time series. In: D. Maidment (ed.), *Handbook of hydrology*, Chapter 17. New York, McGraw-Hill.

SALAS, J.D.; DELLEUR, J.W.; YEJEVICH, V. and LANE, W.L. 1980. *Applied modelling of hydrological time series*. Littleton, Colo., Water Resources Press Publications.

SALAS, J.D. and FERNANDEZ, B. 1993. Models for data generation in hydrology: univariate techniques. In: J.B. Marco, R. Harboe and J.D. Salas (eds), *Stochastic hydrology and its use in water resources systems simulation and optimization*, Dordrecht, Kluwer Academic. pp. 76–95.

SALAS, J.D. and OBEYSEKERA, J.T.B. 1992. Conceptual basis of seasonal streamflow time series. *Journal of Hydraulic Engineering*, Vol. 118, No. 8, pp. 1186–94.

SCHAAKE, J.C. and VICENS, G.J. 1980. Design length of water resource simulation experiments. *Journal of Water Resources Planning and Management*, Vol. 106, No. 1, pp. 333–50.

SLACK, J.R.; WALLIS, J.R. and MATALAS, N.C. 1975. On the value of information in flood frequency analysis. *Water Resources Research*, Vol. 11, No. 5, pp. 629–48.

SMITH, R.L. 1984. Threshold methods for sample extremes. In: J. Tiago de Oliveira (ed.), *Statistical extremes and applications*, Dordrecht, D. Reidel. pp. 621–38.

STEDINGER, J.R. 1980. Fitting log normal distributions to hydrological data. *Water Resources Research*, Vol. 16, No. 3, pp. 481–90.

STEDINGER, J.R. 1981. Estimating correlations in multivariate streamflow models. *Water Resources Research*, Vol. 17, No. 1, pp. 200–08.

STEDINGER, J.R. 1983. Estimating a regional flood frequency distribution. *Water Resources Research*, Vol. 19, No. 2, pp. 503–10.

STEDINGER, J.R. 1997. Expected probability and annual damage estimators. *Journal of Water Resources Planning and Management*, Vol. 123, No. 2, pp. 125–35. [With discussion, Leo R. Beard, 1998. *Journal of Water Resources Planning and Management*, Vol. 124, No. 6, pp. 365–66.]

STEDINGER, J.R. 2000. Flood frequency analysis and statistical estimation of flood risk. In: E.E. Wohl (ed.), *Inland flood hazards: human, riparian and aquatic communities*, Chapter 12. Stanford, UK, Cambridge University Press.

STEDINGER, J.R. and BAKER, V.R. 1987. Surface water hydrology: historical and paleoflood information. *Reviews of Geophysics*, Vol. 25, No. 2, pp. 119–24.

STEDINGER, J.R. and COHN, T.A. 1986. Flood Frequency analysis with historical and paleoflood information. *Water Resources Research*, Vol. 22, No. 5, pp. 785–93.

STEDINGER, J.R. and LU, L. 1995. Appraisal of regional and index flood quantile estimators. *Stochastic Hydrology and Hydraulics*, Vol. 9, No. 1, pp. 49–75.

STEDINGER, J.R.; PEI, D. and COHN, T.A. 1985. A disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations. *Water Resources Research*, Vol. 21, No. 5, pp. 665–7.

STEDINGER, J.R. and TAYLOR, M.R. 1982a. Synthetic streamflow generation. Part 1: model verification and validation. *Water Resources Research*, Vol. 18, No. 4, pp. 919–24.

STEDINGER, J.R. and TAYLOR, M.R. 1982b. Synthetic streamflow generation. Part 2: effect of parameter uncertainty. *Water Resources Research*, Vol. 18, No. 4, pp. 919–24.

STEDINGER, J.R. and VOGEL, R. 1984. Disaggregation procedures for the generation of serially correlated flow vectors. *Water Resources Research*, Vol. 20, No. 1, pp. 47–56.

STEDINGER, J.R.; VOGEL, R.M. and FOUFOULA-GEORGIOU, E. 1993. Frequency analysis of extreme events, In: D. Maidment (ed.), *Handbook of hydrology*, Chapter 18. New York, McGraw-Hill.

STEPHENS, M. 1974. Statistics for Goodness of Fit, *Journal of the American Statistical Association*, Vol. 69, pp. 730–37.

TAO, P.C. and DELLEUR, J.W. 1976. Multistation, multi-year synthesis of hydrological time series by disaggregation. *Water Resources Research*, Vol. 12, No. 6, pp. 1303–11.

TARBOTON, D.G.; SHARMA, A. and LALL, U. 1998. Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water Resources Research*, Vol. 34, No. 1, pp. 107–19.

TASKER, G.D. and STEDINGER, J.R. 1986. Estimating generalized skew with weighted least squares regression. *Journal of Water Resources Planning and Management*, Vol. 112, No. 2, pp. 225–37.

THOM, H.C.S. 1958. A note on the gamma distribution. *Monthly Weather Review*, Vol. 86, No. 4, 1958. pp. 117–22.

THOMAS, H.A., JR. and FIERING, M.B. 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In: A. Maass, M.M. Hufschmidt, R. Dorfman, H.A. Thomas, Jr., S.A. Marglin and G.M. Fair (eds), *Design of water resources systems*. Cambridge, Mass., Harvard University Press.

THYER, M. and KUCZERA, G. 2000. Modelling long-term persistence in hydroclimatic time series using a hidden state Markov model. *Water Resources Research*, Vol. 36, No. 11, pp. 3301–10.

VALDES, J.R.; RODRIGUEZ, I. and VICENS, G. 1977. Bayesian generation of synthetic streamflows 2: the multivariate case. *Water Resources Research*, Vol. 13, No. 2, pp. 291–95.

VALENCIA, R. and SCHAAKE, J.C., Jr. 1973. Disaggregation processes in stochastic hydrology. *Water Resources Research*, Vol. 9, No. 3, pp. 580–5.

VICENS, G.J.; RODRÍGUEZ-ITURBE, I. and SCHAAKE, J.C., Jr. 1975. A Bayesian framework for the use of regional information in hydrology. *Water Resources Research,* Vol. 11, No. 3, pp. 405–14.

VOGEL, R.M. 1987. The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses. *Water Resources Research,* Vol. 22, No. 4, pp. 587–90.

VOGEL, R.M. and FENNESSEY, N.M. 1993. L-moment diagrams should replace product moment diagrams. *Water Resources Research*, Vol. 29, No. 6, pp. 1745–52.

VOGEL, R.M. and MCMARTIN, D.E. 1991. Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type III distribution. *Water Resources Research*, Vol. 27, No. 12, pp. 3149–58.

VOGEL, R.M. and SHALLCROSS, A.L. 1996. The moving blocks bootstrap versus parametric time series models. *Water Resources Research*, Vol. 32, No. 6, pp. 1875–82.

VOGEL, R.M. and STEDINGER, J.R. 1988. The value of stochastic streamflow models in over-year reservoir design applications. *Water Resources Research*, Vol. 24, No. 9, pp. 1483–90.

WALLIS, J.R. 1980. Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures. In: E. Guggino, G. Rossi and E. Todini (eds), *Piene e Siccita,* Catania, Italy, Fondazione Politecnica del Mediterraneo. pp. 3–36.

WALLIS, J.R.; MATALAS, N.C. and SLACK, J.R. 1974a. Just a Moment! *Water Resources Research*, Vol. 10, No. 2, pp. 211–19.

WALLIS, J.R.; MATALAS, N.C. and SLACK, J.R. 1974b. *Just a moment! Appendix*. Springfield, Va., National Technical Information Service, PB-231 816.

WANG, Q. 2001. A. Bayesian joint probability approach for flood record augmentation. *Water Resources Research*, Vol. 37, No. 6, pp. 1707–12.

WANG, Q.J. 1997. LH moments for statistical analysis of extreme events. *Water Resources Research*, Vol. 33, No. 12, pp. 2841–48.

WILK, M.B. and GNANADESIKAN, R. 1968. Probability plotting methods for the analysis of data. *Biometrika*, Vol. 55, No. 1, pp. 1–17.

WILKS, D.S. 2002. Realizations of daily weather in forecast seasonal climate. *Journal of Hydrometeorology*, No. 3, pp. 195–207.

WILKS, D.S. 1998. Multi-site generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, No. 210, pp. 178–91.

YOUNG, G.K. 1968. Discussion of 'Mathematical assessment of synthetic hydrology' by N.C. Matalas and reply. *Water Resources Research*, Vol. 4, No. 3, pp. 681–3.

ZELLNER, A. 1971. An introduction to Bayesian inference in econometrics*,* New York, Wiley.

ZRINJI, Z. and BURN, D.H. 1994. Flood Frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology,* No. 13, pp. 1–21.

# 8. Modelling Uncertainty

1. Introduction 231

2. Generating Values From Known Probability Distributions 231

3. Monte Carlo Simulation 233

4. Chance Constrained Models 235

5. Markov Processes and Transition Probabilities 236

6. Stochastic Optimization 239
   - 6.1. Probabilities of Decisions 243
   - 6.2. A Numerical Example 244

7. Conclusions 251

8. References 251

# 8 Modelling Uncertainty

**Decision-makers are increasingly willing to consider the uncertainty associated with model predictions of the impacts of their possible decisions. Information on uncertainty does not make decision-making easier, but to ignore it is to ignore reality. Incorporating what is known about the uncertainty into input parameters and variables used in optimization and simulation models can help in quantifying the uncertainty in the resulting model predictions – the model output. This chapter discusses and illustrates some approaches for doing this.**

## 1. Introduction

Water resources planners and managers work in an environment of change and uncertainty. Water supplies are always uncertain, if not in the short term at least in the long term. Water demands and the multiple purposes and services water provide are always changing, and these changes cannot always be predicted. Many of the values of parameters of models used to predict the multiple hydrological, economic, environmental, ecological and social impacts are also changing and uncertain. Indeed models used to predict these impacts are, at least in part, based on many imprecise assumptions. Planning and managing, given this uncertainty, cannot be avoided.

To the extent that probabilities can be assigned to various uncertain inputs or parameter values, some of this uncertainty can be incorporated into models. These models are called *probabilistic* or *stochastic* models. Most probabilistic models provide a range of possible values for each output variable along with their probabilities. Stochastic models attempt to model the random processes that occur over time, and provide alternative time series of outputs along with their probabilities. In other cases, sensitivity analyses (solving models under different assumptions) can be carried out to estimate the impact of any uncertainty on the decisions being considered. In some situations, uncertainty may not significantly affect the decisions that should be made. In other situations it will. Sensitivity analyses can help estimate the extent to which we need to try to reduce that uncertainty. Model sensitivity and uncertainty analysis is discussed in more detail in Chapter 9.

This chapter introduces a number of approaches to probabilistic optimization and simulation modelling. Probabilistic models will be developed and applied to some of the same water resources management problems used to illustrate deterministic modelling in previous chapters. They can be, and have been, applied to numerous other water resources planning and management problems as well. The purpose here, however, is simply to illustrate some of the commonly used approaches to the probabilistic modelling of water resources system design and operating problems.

## 2. Generating Values From Known Probability Distributions

Variables whose values cannot be predicted with certainty are called random variables. Often, inputs to hydrological simulation models are observed or synthetically generated values of rainfall or streamflow. Other examples of such random variables could be evaporation losses, point and non-point source wastewater discharges, demands for water, spot prices for energy that may impact the amount of hydropower production, and so on. For random processes that are stationary – that is, the statistical attributes of the process are not changing – and if there is

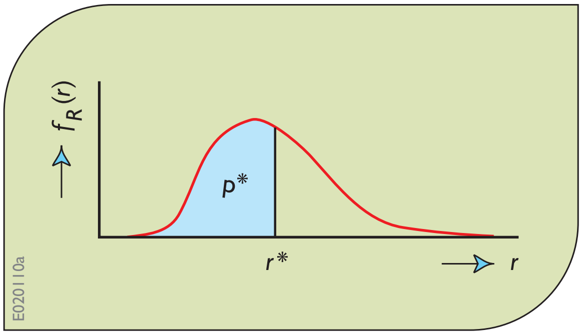



**Figure 8.1.** Probability density distribution of a random variable *R*. The probability that *R* is less than or equal *r** is *p**.

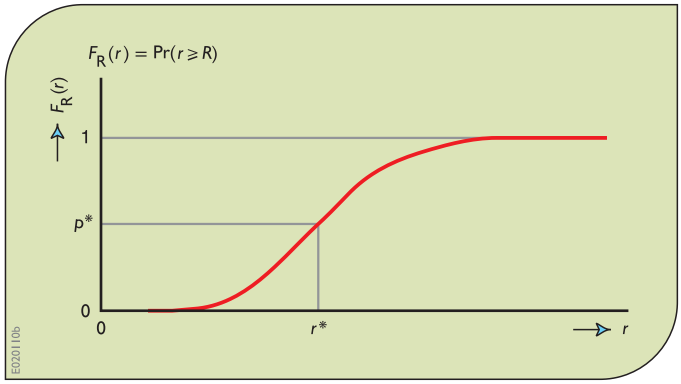



**Figure 8.2.** Cumulative distribution function of a random variable *R* showing the probability of any observed value of *R* being less than or equal to a given value *r*. The probability of an observed value of *R* being less than or equal to *r** is *p**.

no serial correlation in the spatial or temporal sequence of observed values, then such random processes can be characterized by single probability distributions. These probability distributions are often based on past observations of the random variables. These observations or measurements are used either to define the probability distribution itself or to estimate parameter values of an assumed type of distribution.

Let $R$ be a random variable whose probability density distribution, $f_R(r)$, is as shown in Figure 8.1. This distribution indicates the probability or likelihood of an observed value of the random variable $R$ being between any two values of $r$ on the horizontal axis. For example, the probability of an observed value of $R$ being between 0 and $r^*$ is $p^*$, the shaded area to the left of $r^*$. The entire shaded area of a probability density distribution, such as shown in Figure 8.1, is 1.

Integrating this function over $r$ converts the density function to a cumulative distribution function, $F_R(r^*)$, ranging from 0 to 1, as illustrated in Figure 8.2.

$$\int_0^{r^*} f_R(r)\,\mathrm{d}r = \Pr(R \leq r^*) = F_R(r^*) \tag{8.1}$$

Given any value of $p^*$ from 0 to 1, one can find its corresponding variable value $r^*$ from the inverse of the cumulative distribution function.

$$F_R^{-1}(p^*) = r^* \tag{8.2}$$

From the distribution shown in Figure 8.1, it is obvious that the likelihood of different values of the random variable varies; ones in the vicinity of $r^*$ are much more likely to occur than are values at the tails of the distribution. A uniform distribution is one that looks like a rectangle; any value of the random variable between its lower and upper limits is equally likely. Using Equation 8.2, together with a series of uniformly distributed (all equally likely) values of $p^*$ over the range from 0 to 1 (that is, along the vertical axis of Figure 8.2), one can generate a corresponding series of variable values, $r^*$, associated with any distribution. These random variable values will have a cumulative distribution as shown in Figure 8.2, and hence a density distribution as shown in Figure 8.1, regardless of the types or shapes of those distributions. The mean, variance and other moments of the distributions will be maintained.

The mean and variance of continuous distributions are:

$$\int r\, f_R(r)\mathrm{d}r = E[R] \tag{8.3}$$

$$\int (r - E[R])^2\, f_R(r)\mathrm{d}r = \mathrm{Var}[R] \tag{8.4}$$

The mean and variance of discrete distributions having possible values denoted by $r_i$ with probabilities $p_i$ are:

$$\sum_i r_i p_i = E[R] \tag{8.5}$$

$$\sum_i (r_i - E[R])^2 p_i = \mathrm{Var}[R] \tag{8.6}$$

If a time series of $T$ random variable values, $r_t$, from the same stationary random variable, $R$, exists, then the serial or autocorrelations of $r_t$ and $r_{t+k}$ in this time series for any positive integer $k$ can be estimated using:

$$\hat{\rho}_R(k) = \sum_{\tau=1\ T-k} [(r_\tau - E[R])(r_{\tau+k} - E[R])] \Big/ \sum_{t=1\ T} (r_t - E[R])^2 \tag{8.7}$$

The probability density and corresponding cumulative probability distributions can be of any shape, not just those named distributions commonly found in probability and statistics books.

The process of generating a time sequence $t = 1, 2, \ldots$ of inputs, $r_t$, from the probability distribution of a random variable $R$ where the lag 1 serial correlation, $\rho_R(1) = \rho$, is to be preserved is a little more complex. The expected value of the random variable $R_{t+1}$ depends on the observed value, $r_t$, of the random variable $R_t$, together with the mean of the distribution, $E[R]$, and the correlation coefficient $\rho$. If there is no correlation ($\rho$ is 0), then the expected value of $R_{t+1}$ is the mean of the population, $E[R]$. If there is perfect correlation ($\rho$ is 1), then the expected value of $R_{t+1}$ is $r_t$. In general, the expected value of $R_{t+1}$ given an observed value $r_t$ of $R_t$ is:

$$E[R_{t+1}|R_t = r_t] = E[R] + \rho(r_t - E[R]) \tag{8.8}$$

The variance of the random variable $R_{t+1}$ depends on the variance of the distribution, $\text{Var}[R]$, and the lag 1 correlation coefficient, $\rho$.

$$\text{Var}[R_{t+1}|R_t = r_t] = \text{Var}[R](1 - \rho^2) \tag{8.9}$$

If there is perfect correlation ($\rho = 1$), then the process is deterministic, there is no variance, and $r_{t+1} = r_t$. The value for $r_{t+1}$ is $r_t$. If there is no correlation – that is, serial correlation does not exist ($\rho = 0$) – then the generated value for $r_{t+1}$ is its mean, $E[R]$, plus some randomly generated deviation from a normal distribution having a mean of 0 and a standard deviation of 1, denoted as $N(0, 1)$. In this case the value $r_{t+1}$ is not dependent on $r_t$.

When the serial correlation is more than 0 but less than 1, then both the correlation and the standard deviation (the square root of the variance) influence the value of $r_{t+1}$. A sequence of random variable values from a multivariate normal distribution that preserves the mean, $E[R]$; overall variance, $\text{Var}[R]$; standard deviation $\sigma$, and lag 1 correlation $\rho$; can be obtained from.

$$r_{t+1} = E[R] + \rho(r_t - E[R]) + Z\,\sigma(1 - \rho^2)^{1/2} \tag{8.10}$$

The term $Z$ in Equation 8.10 is a random number generated from a normal distribution having a mean of 0 and a variance of 1. The process involves selecting a random number from a uniform distribution ranging from 0 to 1, and using it in Equation 8.2 for an $N(0, 1)$ distribution to obtain a value of random number for use in Equation 8.10. This positive or negative number is substituted for the term $Z$ in Equation 8.10 to obtain a value $r_{t+1}$. This is shown on the graph in Figure 8.3.

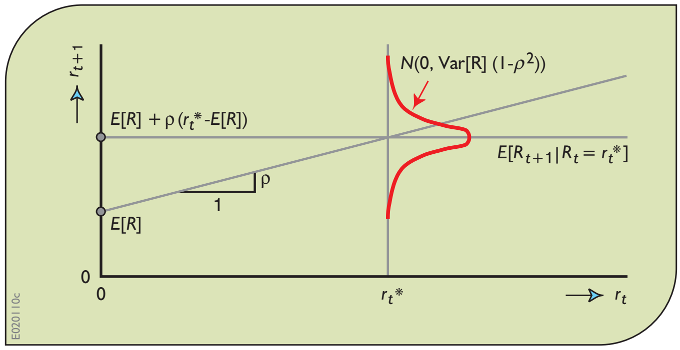



**Figure 8.3.** Diagram showing the calculation of a sequence of values of the random variable $R$ from a multivariate normal distribution in a way that preserves the mean, variance and correlation of the random variable.

Simulation models that have random inputs, such as a series of $r_t$ values, will generally produce random outputs. After many simulations, the probability distributions of each random output variable value can be defined. These can be used to estimate reliabilities and other statistical characteristics of those output distributions. This process of generating multiple random inputs for multiple simulations to obtain multiple random outputs is called Monte Carlo simulation.

## 3. Monte Carlo Simulation

To illustrate Monte Carlo simulation, consider the water allocation problem involving three firms, each of which receives a benefit, $B_i(x_{it})$, from the amount of water, $x_{it}$, allocated to it in each period $t$. This situation is shown in Figure 8.4. Monte Carlo simulation can be used to find the probability distribution of the benefits to each firm associated with the firm's allocation policy.

Suppose the policy is to keep the first two units of flow in the stream, to allocate the next three units to Firm 3, and the next four units to Firms 1 and 2 equally. The remaining flow is to be allocated to each of the three firms equally up to the limits desired by each firm, namely 3.0, 2.33, and 8.0 respectively. Any excess flow will remain in the stream. The plots in Figure 8.5 illustrate this policy. Each allocation plot reflects the priorities given to the three firms and the users further downstream.

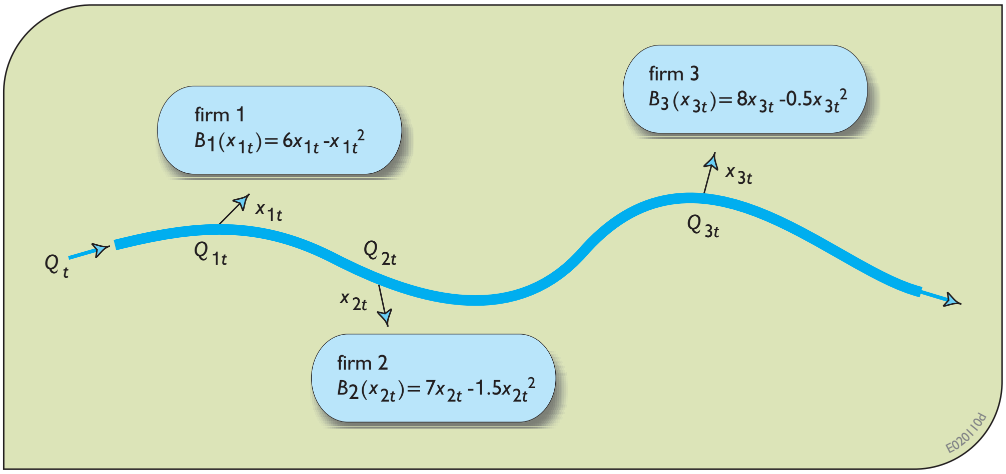



**Figure 8.4.** Streamflow allocations in each period $t$ result in benefits, $B_i(x_{it})$, to each firm $i$. The flows, $Q_{it}$, at each diversion site $i$ are the random flows $Q_t$ less the upstream withdrawals, if any.

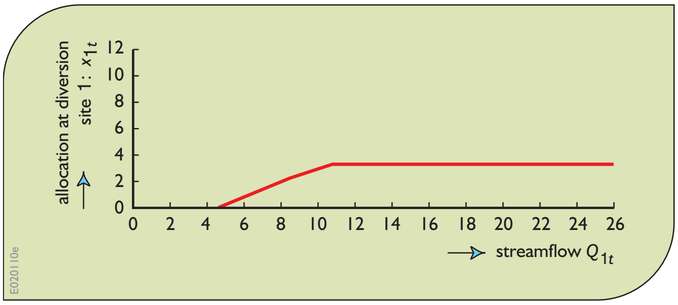



**Figure 8.5a.** Water allocation policy for Firm 1 based on the flow at its diversion site. This policy applies for each period $t$.

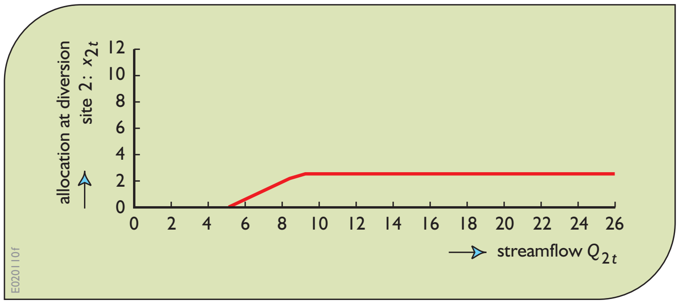



**Figure 8.5b.** Water allocation policy for Firm 2 based on the flow at its diversion site for that firm. This policy applies for each period $t$.

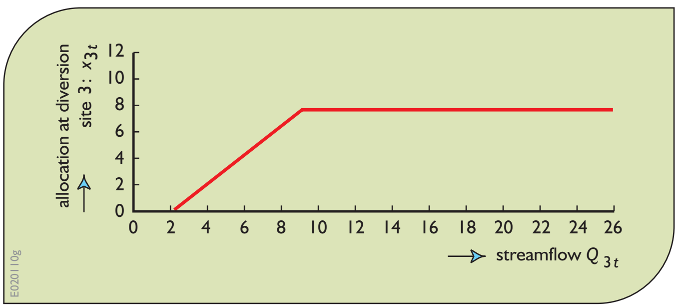



**Figure 8.5c.** Water allocation policy for Firm 3 based on the flow at its diversion site. This policy applies for each period $t$.

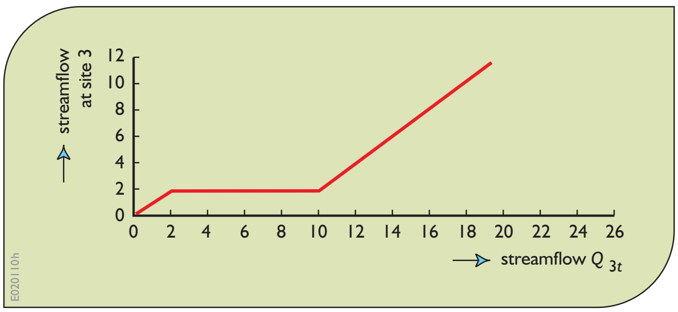



**Figure 8.5d.** Streamflow downstream of site 3 given the streamflow $Q_{3t}$ at site 3 before the diversion. This applies for each period $t$.

A simulation model can be created. In each of a series of discrete time periods $t$, the flows $Q_t$ are drawn from a probability distribution, such as from Figure 8.2 using Equation 8.2. Once this flow is determined, each successive allocation, $x_{it}$, is computed. Once an allocation is made it is subtracted from the streamflow and the next allocation is made on the basis of that reduced streamflow, in accordance with the allocation policy defined in Figures 8.5a – d. After numerous time steps, the probability distributions of the allocations to each of the firms can be defined.

Figure 8.6 shows a flow chart for this simulation model.

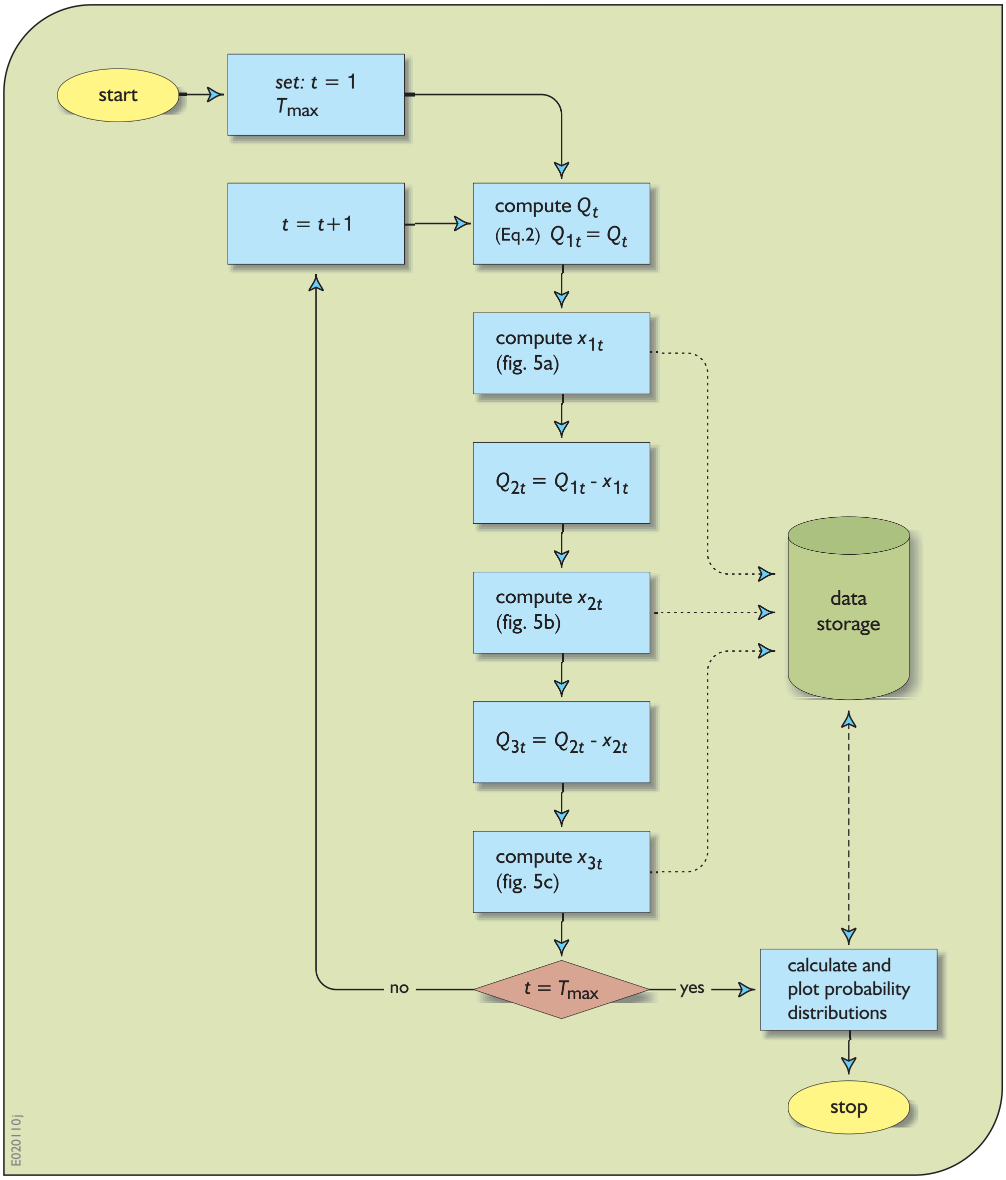



**Figure 8.6.** Monte Carlo simulation to determine probability distributions of allocations to each of three water users, as illustrated in Figure 8.4. The dashed lines represent information (data) flows.

Having defined the probability distribution of the allocations, based on the allocation policy, one can consider each of the allocations as random variables, $X_1$, $X_2$, and $X_3$ for Firms 1, 2 and 3 respectively.

## 4. Chance Constrained Models

For models that include random variables, it may be appropriate in some situations to consider constraints that do not have to be satisfied all the time. Chance constraints specify the probability of a constraint being satisfied, or the fraction of the time a constraint has to apply. Consider, for example, the allocation problem shown in Figure 8.4. For planning purposes, the three firms may want to set allocation targets, not expecting to have those targets met 100% of the time. To ensure, for example, that an allocation target, $T_i$, of firm $i$ will be met at least 90% of the time, one could write the chance constraint:

$$\Pr\{T_i \leq X_i\} \geq 0.90 \quad i = 1, 2 \text{ and } 3 \tag{8.11}$$

In this constraint, the allocation target $T_i$ is an unknown decision-variable, and $X_i$ is a random variable whose distribution has just been computed and is known.

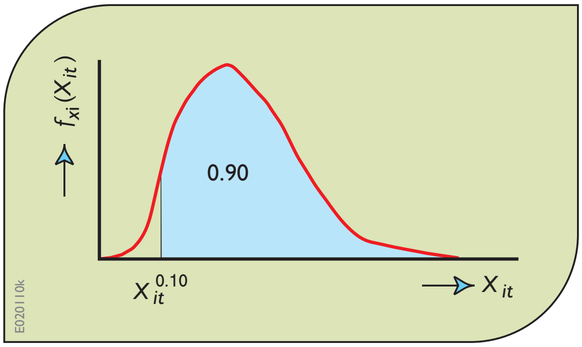



**Figure 8.7.** Probability density distribution of the random allocation $X_i$ to firm $i$. The particular allocation value $x_{it}^{0.10}$ has a 90% chance of being equalled or exceeded, as indicated by the shaded region.

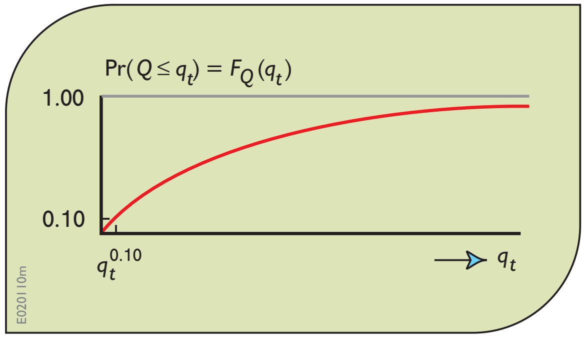



**Figure 8.8.** Example cumulative probability distribution showing the particular value of the random variable, $q_t^{0.10}$, that is equalled or exceeded 90% of the time.

To include chance constraints in optimization models, their deterministic equivalents must be defined. The deterministic equivalents of the three chance constraints in Equation 8.11 are:

$$T_i \le x_{it}^{0.10} \quad i = 1, 2 \text{ and } 3 \tag{8.12}$$

where $x_{it}^{0.10}$ is the particular value of the random variable $X_i$ that is equalled or exceeded 90% of the time. This value is shown on the probability distribution for $X_i$ in Figure 8.7.

To modify the allocation problem somewhat, assume the benefit obtained by each firm is a function of its target allocation and that the same allocation target applies in each time period $t$. The equipment and labour used in the firm is presumably based on the target allocations. Once the target is set, assume there are no benefits gained by excess water allocations. If the benefits obtained are to be based on the target allocations rather than the actual allocations, then the optimization problem is one of finding the values of the three targets that maximize the total benefits obtained with a reliability of, say, at least 90%.

$$\text{Maximize}\left(6T_1 - T_1^2\right) + \left(7T_2 - 1.5T_2^2\right) + \left(8T_3 - 0.5T_3^2\right) \tag{8.13}$$

subject to:

$$\Pr\{T_1 + T_2 + T_3 \le [Q_t - \min(Q_t, 2)]\} \ge 0.90 \quad \text{for all periods } t \tag{8.14}$$

where $Q_t$ is the random streamflow variable upstream of all diversion sites. If the same unconditional probability distribution of $Q_t$ applies for each period $t$, then only one Equation 8.14 is needed.

Assuming the value of the streamflow, $q_t^{0.10}$, that is equalled or exceeded 90% of the time, is greater than 2 (the amount that must remain in the stream), the deterministic equivalent of chance constraint Equation 8.14 is:

$$T_1 + T_2 + T_3 \le \left[q_t^{0.10} - \min\left(q_t^{0.10}, 2\right)\right] \tag{8.15}$$

The value of the flow that is equal to or exceeds 90% of the time, $q_t^{0.10}$, can be obtained from the cumulative distribution of flows as illustrated in Figure 8.8.

Assume this 90% reliable flow is 8. The deterministic equivalent of the chance constraint Equation 8.9 for all periods $t$ is simply $T_1 + T_2 + T_3 \le 6$. The optimal solution of the chance-constrained target allocation model, Equations 8.8 and 8.9, is, as seen before, $T_1 = 1$, $T_2 = 1$ and $T_3 = 4$. The next step would be to simulate this problem to see what the actual reliabilities might be for various sequences of flows $q_t$.

## 5. Markov Processes and Transition Probabilities

Time-series correlations can be incorporated into models using transition probabilities. To illustrate this process, consider the observed flow sequence shown in Table 8.1.

The estimated mean, variance and correlation coefficient of the observed flows shown in Table 8.1 can be calculated using Equations 8.16, 8.17 and 8.18.

| period $t$ | flow $Q_t$ | period $t$ | flow $Q_t$ | period $t$ | flow $Q_t$ |
|---|---|---|---|---|---|
| 1 | 4.5 | 11 | 1.8 | 21 | 1.8 |
| 2 | 5.2 | 12 | 2.5 | 22 | 1.2 |
| 3 | 6.0 | 13 | 2.3 | 23 | 2.5 |
| 4 | 3.2 | 14 | 1.8 | 24 | 1.9 |
| 5 | 4.3 | 15 | 1.2 | 25 | 3.2 |
| 6 | 5.1 | 16 | 1.9 | 26 | 2.5 |
| 7 | 3.6 | 17 | 2.5 | 27 | 3.5 |
| 8 | 4.5 | 18 | 4.1 | 28 | 2.7 |
| 9 | 1.8 | 19 | 4.7 | 29 | 1.5 |
| 10 | 1.5 | 20 | 5.6 | 30 | 4.1 |
| | | | | 31 | 4.8 |

E020827w

**Table 8.1.** Sequence of flows for thirty-one time periods *t*.

$$E[Q] = \sum_{1}^{31} q_t/31 = 3.155 \tag{8.16}$$

$$\text{Var}[Q] = \sum_{1}^{31} (q_t - 3.155)^2/30 = 1.95 \tag{8.17}$$

Lag-one correlation coefficient $= \rho$

$$= \left[\sum_{1}^{30} (q_{t+1} - 3.155)(q_t - 3.155)\right] \Bigg/ \sum_{1}^{31} (q_t - 3.155)^2 = 0.50 \tag{8.18}$$

The probability distribution of the flows in Table 8.1 can be approximated by a histogram. Histograms can be created by subdividing the entire range of random variable values, such as flows, into discrete intervals. For example, let each interval be two units of flow. Counting the number of flows in each interval and then dividing those interval counts by the total number of counts results in the histogram shown in Figure 8.9. In this case, just to compare this with what will be calculated later, the first flow, $q_1$, is ignored.

Figure 8.9 shows a uniform unconditional probability distribution of the flow being in any of the possible discrete flow intervals. It does not show the possible dependency of the probabilities of the random variable value, $q_{t+1}$, in period $t + 1$ on the observed random variable value, $q_t$, in period $t$. It is possible that the probability of being in a flow interval $j$ in period $t + 1$ depends on the actual observed flow interval $i$ in period $t$.

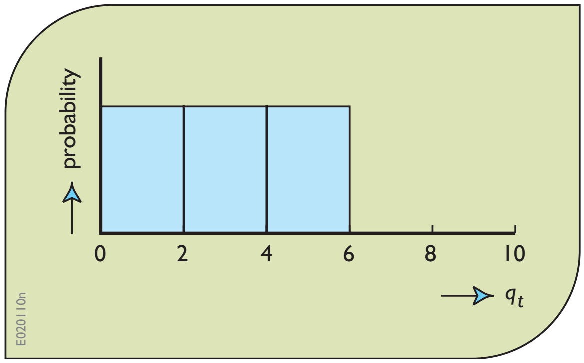



**Figure 8.9.** Histogram showing an equal 1/3 probability that the values of the random variable $Q_t$ will be in any one of the three two-flow unit intervals.

| flow interval in t: $i$ | flow interval in t + 1: $j$ = 1 | 2 | 3 |
|---|---|---|---|
| 1 | 5 | 4 | 1 |
| 2 | 3 | 4 | 3 |
| 3 | 2 | 2 | 6 |

E020827x

**Table 8.2.** Matrix showing the number of times a flow in interval *i* in period *t* was followed by a flow in interval *j* in period $t + 1$.

To see if the probability of being in any given interval of flows is dependent on the past flow interval, one can create a matrix. The rows of the matrix are the flow intervals $i$ in period $t$. The columns are the flow intervals $j$ in the following period $t + 1$. Such a matrix is shown in Table 8.2. The numbers in the matrix are based on the flows in Table 8.1 and indicate the number of times a flow in interval $j$ followed a flow in interval $i$.

Given an observed flow in an interval $i$ in period $t$, the probabilities of being in one of the possible intervals $j$ in the next period $t + 1$ must sum to 1. Thus, each number in each row of the matrix in Table 8.2 can be divided by the total number of flow transitions in that row (the sum

| flow interval in t: $i$ | flow interval in t + 1: $j$ 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0.5 | 0.4 | 0.1 |
| 2 | 0.3 | 0.4 | 0.3 |
| 3 | 0.2 | 0.2 | 0.6 |

E020911b

**Table 8.3.** Matrix showing the probabilities $P_{ij}$ of having a flow in interval $j$ in period $t+1$ given an observed flow in interval $i$ in period $t$.

of the number of flows in the row) to obtain the probabilities of being in each interval $j$ in $t+1$ given a flow in interval $i$ in period $t$. In this case there are ten flows that followed each flow interval $i$, hence by dividing each number in each row of the matrix by 10 defines the transition probabilities $P_{ij}$.

$$P_{ij} = \Pr\{Q_{t+1} \text{ in interval } j \mid Q_t \text{ in interval } i\} \qquad (8.19)$$

These conditional or transition probabilities, shown in Table 8.3, correspond to the number of transitions shown in Table 8.2.

Table 8.3 is a matrix of transition probabilities. The sum of the probabilities in each row equals 1. Matrices of transition probabilities whose rows sum to 1 are also called stochastic matrices or first-order Markov chains.

If each row's probabilities were the same, this would indicate that the probability of observing any flow interval in the future is independent of the value of previous flows. Each row would have the same probabilities as the unconditional distribution shown in Figure 8.9. In this example the probabilities in each row differ, showing that low flows are more likely to follow low flows, and high flows are more likely to follow high flows. Thus the flows in Table 8.1 are positively correlated, as indeed has already determined from Equation 8.18.

Using the information in Table 8.3, one can compute the probability of observing a flow in any interval at any period on into the future given the present flow interval. This can be done one period at a time. For example assume the flow in the current time period $t = 1$ is in interval $i = 3$. The probabilities, $PQ_{j,2}$, of being in any of the three intervals in the following time period $t = 2$ are the probabilities shown in the third row of the matrix in Table 8.3.

The probabilities of being in an interval $j$ in the following time period $t = 3$ is the sum over all intervals $i$ of the joint probabilities of being in interval $i$ in period $t = 2$ and making a transition to interval $j$ in period $t = 3$.

$$\Pr\{Q_3 \text{ in interval } j\} = PQ_{j,3}$$
$$= \sum_i \Pr\{Q_2 \text{ in interval } i\}$$
$$\Pr\{Q_3 \text{ in interval } j \mid Q_2 \text{ in interval } i\} \qquad (8.20)$$

The last term in Equation 8.20 is the transition probability, from Table 8.3, that in this example remains the same for all time periods $t$. These transition probabilities, $\Pr\{Q_{t+1} \text{ in interval } j \mid Q_t \text{ in interval } i\}$ can be denoted as $P_{ij}$.

Referring to Equation 8.19, Equation 8.20 can be written in a general form as:

$$PQ_{j,t+1} = \sum_i PQ_{it} P_{ij} \text{ for all intervals } j \text{ and periods } t \qquad (8.21)$$

This operation can be continued to any future time period. Table 8.4 illustrates the results of such calculations for

| time period $t$ | flow interval $i$ 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | 0 | 1 |
| 2 | 0.2 | 0.2 | 0.6 |
| 3 | 0.28 | 0.28 | 0.44 |
| 4 | 0.312 | 0.312 | 0.376 |
| 5 | 0.325 | 0.325 | 0.350 |
| 6 | 0.330 | 0.330 | 0.340 |
| 7 | 0.332 | 0.332 | 0.336 |
| 8 | 0.333 | 0.333 | 0.334 |
| | probability $PQ_{it}$ | | |

E020827y

**Table 8.4.** Probabilities of observing a flow in any flow interval $i$ in a future time period $t$ given a current flow in interval $i = 3$. These probabilities are derived using the transition probabilities $P_{ij}$ in Table 8.3 in Equation 8.21 and assuming the flow interval observed in Period 1 is in Interval 3.

up to six future periods, given a present period ($t = 1$) flow in interval $i = 3$.

Note that as the future time period $t$ increases, the flow interval probabilities are converging to the unconditional probabilities – in this example 1/3, 1/3, 1/3 – as shown in Figure 8.9. The predicted probability of observing a future flow in any particular interval at some time in the future becomes less and less dependent on the current flow interval as the number of time periods increases between the current period and that future time period.

When these unconditional probabilities are reached, $PQ_{it}$ will equal $PQ_{i,t+1}$ for each flow interval $i$. To find these unconditional probabilities directly, Equation 8.21 can be written as:

$$PQ_j = \sum_i PQ_i P_{ij} \quad \text{for all intervals } j \tag{8.22}$$

Equation 8.22 (less one) along with Equation 8.23 can be used to calculate all the unconditional probabilities $PQ_i$ directly.

$$\sum_i PQ_i = 1 \tag{8.23}$$

Conditional or transition probabilities can be incorporated into stochastic optimization models of water resources systems.

## 6. Stochastic Optimization

To illustrate the development and use of stochastic optimization models, consider first the allocation of water to a single user. Assume the flow in the stream where the diversion takes place is not regulated and can be described by a known probability distribution based on historical records. Clearly, the user cannot divert more water than is available in the stream. A deterministic model would include the constraint that the diversion $x$ cannot exceed the available water $Q$. But $Q$ is a random variable. Some target value, $q$, of the random variable $Q$ will have to be selected, knowing that there is some probability that in reality, or in a simulation model, the actual flow may be less than the selected value $q$. Hence, if the constraint $x \leq q$ is binding, the actual allocation may be less than the value of the allocation or diversion variable $x$ produced by the optimization model.

If the value of $x$ affects one of the system's performance indicators, such as the net benefits, $B(x)$, to the user, a more accurate estimate of the user's net benefits will be obtained from considering a range of possible allocations $x$, depending on the range of possible values of the random flow $Q$. One way to do this is to divide the known probability distribution of flows $q$ into discrete ranges, $i$ – each range having a known probability $PQ_i$. Designate a discrete flow $q_i$ for each range. Associated with each specified flow $q_i$ is an unknown allocation $x_i$. Now the single deterministic constraint $x \leq q$ can be replaced with the set of deterministic constraints $x_i \leq q_i$, and the term $B(x)$ in the original objective function can be replaced by its expected value, $\sum_i PQ_i \cdot B(x_i)$.

Note, when dividing a continuous known probability distribution into discrete ranges, the discrete flows $q_i$, selected to represent each range $i$ having a given probability $PQ_i$, should be selected so as to maintain at least the mean and variance of that known distribution as defined by Equations 8.5 and 8.6.

To illustrate this, consider a slightly more complex example involving the allocation of water to consumers upstream and downstream of a reservoir. Both the policies for allocating water to each user and the reservoir release policy are to be determined. This example problem is shown in Figure 8.10.

If the allocation of water to each user is to be based on a common objective, such as the minimization of the total sum, over time, of squared deviations from pre-specified target allocations, each allocation in each time period will depend in part on the reservoir storage volume.

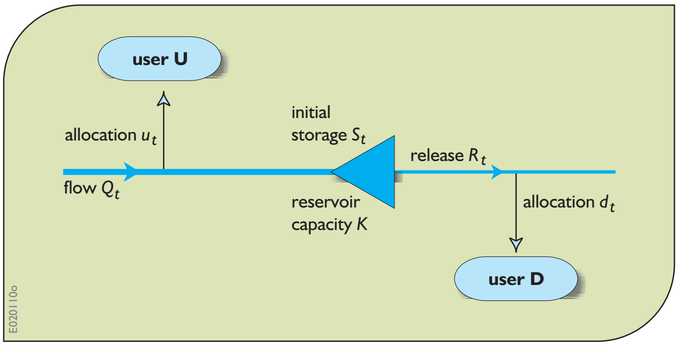



**Figure 8.10.** Example water resources system involving water diversions from a river both upstream and downstream of a reservoir of known capacity.

Consider first a deterministic model of the above problem, assuming known river flows $Q_t$ and upstream and downstream user allocation targets $UT_t$ and $DT_t$ in each of $T$ within-year periods $t$ in a year. Assume the objective is to minimize the sum of squared deviations from actual allocations, $u_t$ and $d_t$, and their respective target allocations, $UT_t$ and $DT_t$ in each within-year period $t$.

$$\text{Minimize} \sum_t^T \{(UT_t - u_t)^2 + (DT_t - d_t)^2\} \tag{8.24}$$

The constraints include:

a) Continuity of storage involving initial storage volumes $S_t$, net inflows $Q_t - u_t$, and releases $R_t$. Assuming no losses:

$$S_t + Q_t - u_t - R_t = S_{t+1} \quad \text{for each period } t, \; T+1 = 1 \tag{8.25}$$

b) Reservoir capacity limitations. Assuming a known active storage capacity $K$:

$$S_t \le K \quad \text{for each period } t \tag{8.26}$$

c) Allocation restrictions. For each period $t$:

$$u_t \le Q_t \tag{8.27}$$

$$d_t \le R_t \tag{8.28}$$

Equations 8.25 and 8.28 could be combined to eliminate the release variable $R_t$, since in this problem knowledge of the total release in each period $t$ is not required. In this case, Equation 8.25 would become an inequality.

The solution of this model, Equations 8.24 – 8.28, would depend on the known variables (the targets $UT_t$ and $DT_t$, flows $Q_t$ and reservoir capacity $K$). It would identify the particular upstream and downstream allocations and reservoir releases in each period $t$. It would not provide a policy that defines what allocations and releases to make for a range of different inflows and initial storage volumes in each period $t$. A backward-moving dynamic programming model can provide such a policy. This policy will identify the allocations and releases to make based on various initial storage volumes, $S_t$, and flows, $Q_t$, as discussed in Chapter 4.

This deterministic discrete dynamic programming allocation and reservoir operation model can be written for different discrete values of $S_t$ from $0 \le S_t \le$ capacity $K$ as:

$$F_t^n(S_t, Q_t) = \min\{(UT_t - u_t)^2 + (DT_t - d_t)^2 + F_{t+1}^{n-1}(S_{t+1}, Q_{t+1})\}$$

The minimization is over all feasible $u_t$, $R_t$, $d_t$:

$$u_t \le Q_t$$

$$R_t \le S_t + Q_t - u_t$$

$$R_t \ge S_t + Q_t - u_t - K$$

$$d_t \le R_t$$

$$S_{t+1} = S_t + Q_t - u_t - R_t \tag{8.29}$$

There are three variables to be determined at each stage or time period $t$ in the above dynamic programming model. These three variables are the allocations $u_t$ and $d_t$ and the reservoir release $R_t$. Each decision involves three discrete decision-variable values. The functions $F_t^n(S_t, Q_t)$ define the minimum sum of squared deviations given an initial storage volume $S_t$ and streamflow $Q_t$ in time period or season $t$ with $n$ time periods remaining until the end of reservoir operation.

One can reduce this three decision-variable model to a single variable model by realizing that, for any fixed discrete pair of initial and final storage volume states, there can be a direct tradeoff between the upstream and downstream allocations, given the particular streamflow in each period $t$. Increasing the upstream allocation will decrease the resulting reservoir inflow, and this in turn will reduce the release by the same amount. This reduces the amount of water available to allocate to the downstream use.

Hence, for this example problem involving these upstream and downstream allocations, a local optimization can be performed at each time step $t$ for each combination of storage states $S_t$ and $S_{t+1}$. This optimization finds the allocation decision-variables $u_t$ and $d_t$ that

$$\text{minimize}(UT_t - u_t)^2 + (DT_t - d_t)^2 \tag{8.30}$$

where

$$u_t \le Q_t \tag{8.31}$$

$$d_t \le S_t + Q_t - u_t - S_{t+1} \tag{8.32}$$

This local optimization can be solved to identify the $u_t$ and $d_t$ allocations for each feasible combination of $S_t$ and $S_{t+1}$ in each period $t$.

Given these optimal allocations, the dynamic programming model can be simplified to include only one discrete decision-variable, either $R_t$ or $S_{t+1}$. If the decision variable $S_{t+1}$ is used in each period $t$, the releases $R_t$ in those periods $t$ do not need to be considered. Thus the dynamic programming model expressed by Equations 8.29 can be written for all discrete storage volumes $S_t$ from 0 to $K$ and for all discrete flows $Q_t$ as:

$$F_t^n(S_t, Q_t) = \min\{(UT_t - u_t(S_t, S_{t+1}))^2 + (DT_t - d_t(S_t, S_{t+1}))^2 + F_{t+1}^{n-1}(S_{t+1}, Q_{t+1})\}$$

The minimization is over all feasible discrete values of $S_{t+1}$,

$$S_{t+1} \leq K \tag{8.33}$$

where the functions $u_t(S_t, S_{t+1})$ and $d_t(S_t, S_{t+1})$ have been determined using Equations 8.30 – 8.32.

As the total number of periods remaining, $n$, increases, the solution of this dynamic programming model will converge to a steady or stationary state. The best final storage volume $S_{t+1}$ given an initial storage volume $S_t$ will probably differ for each within-year period or season $t$, but for a given season $t$ it will be the same in successive years. In addition, for each storage volume $S_t$, streamflow, $Q_t$, and within-year period $t$, the difference between $F_t^{n+T}(S_t, Q_t)$ and $F_t^n(S_t, Q_t)$ will be the same constant regardless of the storage volume $S_t$ and period $t$. This constant is the optimal, in this case minimum, annual value of the objective function, Equation 8.24.

There could be additional limits imposed on storage variables and release variables, such as for flood control storage or minimum downstream flows, as might be appropriate in specific situations.

The above deterministic dynamic programming model (Equation. 8.33) can be converted to a stochastic model. Stochastic models consider multiple discrete flows as well as multiple discrete storage volumes, and their probabilities, in each period $t$. A common way to do this is to assume that the sequence of flows follow a first-order Markov process. Such a process involves the use of transition or conditional probabilities of flows as defined by Equation 8.20.

To develop these stochastic optimization models, it is convenient to introduce some additional indices or subscripts. Let the index $k$ denote different initial storage volume intervals. These discrete intervals divide the continuous range of storage volume values from 0 to the active reservoir capacity $K$. Each $S_{kt}$ is a discrete storage volume that represents the range of storage volumes in interval $k$ at the beginning of each period $t$.

Let the following letter $l$ be the index denoting different final storage volume intervals. Each $S_{l,t+1}$ is a discrete volume that represents the storage volume interval $l$ at the end of each period $t$ or equivalently at the beginning of period $t + 1$. As previously defined, let the indices $i$ and $j$ denote the different flow intervals, and each discrete flow $q_{it}$ and $q_{j,t+1}$ represent those flow intervals $i$ and $j$ in periods $t$ and $t + 1$ respectively.

These subscripts and the volume or flow intervals they represent are illustrated in Figure 8.11.

With this notation, it is now possible to develop a stochastic dynamic programming model that will identify the allocations and releases that are to be made given both the initial storage volume, $S_{kt}$, and the flow, $q_{it}$. It follows the same structure as the deterministic models defined by Equations 8.30 through 8.32, and 8.33.

To identify the optimal allocations in each period $t$ for each pair of feasible initial and final storage volumes $S_{kt}$ and $S_{l,t+1}$, and inflows $q_{it,}$ one can solve Equations 8.34 through 8.36.

$$\text{minimize } (UT_t - u_{kit})^2 + (DT_t - d_{kilt})^2 \tag{8.34}$$

where

$$u_{kit} \leq q_{it} \quad \forall\ k, i, t. \tag{8.35}$$

$$d_{kilt} \leq S_{kt} + q_{it} - u_{kit} - S_{l,t+1} \quad \forall \text{ feasible } k, i, l, t. \tag{8.36}$$

The solution to these equations for each feasible combination of intervals $k$, $i$, $l$, and period $t$ defines the optimal allocations that can be expressed as $u_t(k, i)$ and $d_t(k, i, l)$.

The stochastic version of Model 8.33, again expressed in a form suitable for backward-moving discrete dynamic programming, can be written for different discrete values of $S_{kt}$ from 0 to $K$ and for all $q_{it}$ as:

$$F_t^n(S_{kt}, q_{it}) = \min\left\{(UT_t - u_t(k,t))^2 + (DT_t - d_t(k,i,l))^2 + \sum_j P_{ij}^t F_{t+1}^{n-1}(S_{1,t+1}, q_{j,t+1})\right\}$$

The minimization is over all feasible discrete values of $S_{l,t+1}$

$$S_{l,t+1} \leq K$$

$$S_{l,t+1} \leq S_{kt} + q_{it} \tag{8.37}$$

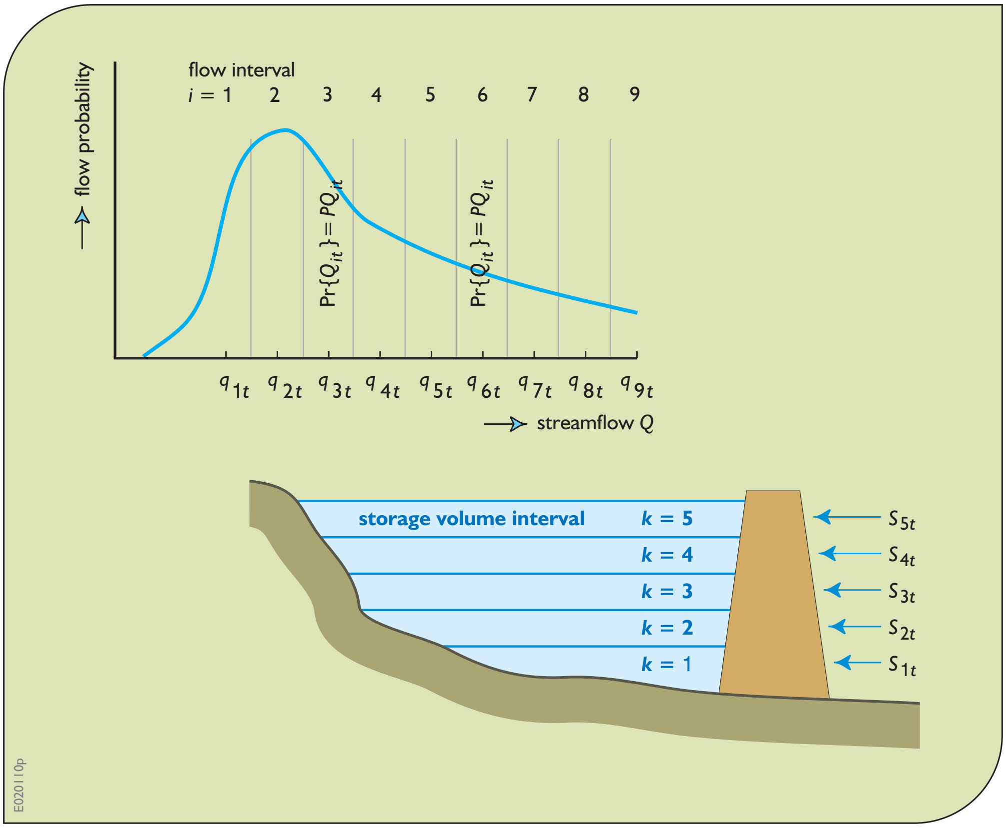



**Figure 8.11.** Discretization of streamflows and reservoir storage volumes. The area within each flow interval *i* below the probability density distribution curve is the unconditional probability, $PQ_{it}$, associated with the discrete flow $q_{it}$.

Each $P^t_{ij}$ in the above recursive equation is the known conditional or transition probability of a flow $q_{j,t+1}$ within interval $j$ in period $t+1$ given a flow of $q_{it}$ within interval $i$ in period $t$.

$$P^t_{ij} = \Pr\{\text{flow } q_{j,t+1} \text{ within interval } j \text{ in } t+1 \mid \text{flow of } q_{it} \text{ within interval } i \text{ in } t\}$$

The sum over all flow intervals $j$ of these conditional probabilities times the $F^{n-1}_{t+1}(S_{l,t+1}, q_{j,t+1})$ values is the expected minimum sum of future squared deviations from allocation targets with $n-1$ periods remaining given an initial storage volume of $S_{kt}$ and flow of $q_{it}$ and final storage volume of $S_{l,t+1}$. The value $F^n_t(S_{kt}, q_{it})$ is the expected minimum sum of squared deviations from the allocation targets with $n$ periods remaining given an initial storage volume of $S_{kt}$ and flow of $q_{it}$. Stochastic models such as these provide expected values of objective functions.

Another way to write the recursion equations of this model, Equation 8.37, is by using just the indices $k$ and $l$ to denote the discrete storage volume variables $S_{kt}$ and $S_{l,t+1}$ and indices $i$ and $j$ to denote the discrete flow variables $q_{it}$ and $q_{j,t+1}$:

$$F^n_t(k,i) = \min_l \left\{ (UT_t - u_t(k,t))^2 + (DT_t - d_t(k,i,l))^2 + \sum_j P^t_{ij} F^{n-1}_{t+1}(l,j) \right\}$$

such that $S_{l,t+1} \le K$

$$S_{l,t+1} \le S_{kt} + q_{it} \tag{8.38}$$

The steady-state solution of this dynamic programming model will identify the preferred final storage volume $S_{l,t+1}$ in period $t$ given the particular discrete initial storage volume $S_{kt}$ and flow $q_{it}$. This optimal policy can be expressed as a function $\ell$ that identifies the best interval $l$ given intervals $k$, $i$ and period $t$.

$$l = \ell(k, i, t) \tag{8.39}$$

All values of $l$ given $k$, $i$ and $t$, defined by Equation 8.39, can be expressed in a matrix, one for each period $t$.

Knowing the best final storage volume interval $l$ given an initial storage volume interval $k$ and flow interval $i$, the optimal downstream allocation, $d_t(k, i)$, can, like the upstream allocation, be expressed in terms of only $k$ and $i$ in each period $t$. Thus, knowing the initial storage volume $S_{kt}$ and flow $q_{it}$ is sufficient to define the optimal allocations $u_t(k, i)$ and $d_t(k, i)$, final storage volume $S_{l,t+1}$, and hence the release $R_t(k, i)$.

$$S_{kt} + q_{it} - u_t(k, i) - R_t(k, i) = S_{l,t+1} \quad \forall\ k, i, t$$
$$\text{where } l = \ell(k, i, t) \tag{8.40}$$

## 6.1. Probabilities of Decisions

Knowing the function $l = \ell(k, i, t)$ permits a calculation of the probabilities of the different discrete storage volumes, allocations, and flows. Let

$PS_{kt}$ = the unknown probability of an initial storage volume $S_{kt}$ being within some interval $k$ in period $t$;

$PQ_{it}$ = the steady-state unconditional probability of flow $q_{it}$ within interval $i$ in period $t$; and

$P_{kit}$ = the unknown probability of the upstream and downstream allocations $u_t(k, i)$ and $d_t(k, i)$ and reservoir release $R_t(k, i)$ in period $t$.

As previously defined,
$P^t_{ij}$ = the known conditional or transition probability of a flow within interval $j$ in period $t + 1$ given a flow within interval $i$ in period $t$.

These transition probabilities $P^t_{ij}$ can be displayed in matrices, similar to Table 8.3, but as a separate matrix (Markov chain) for each period $t$.

The joint probabilities of an initial storage interval $k$, an inflow in the interval $i$, $P_{kit}$ in each period $t$ must satisfy two conditions. Just as the initial storage volume in period $t + 1$ is the same as the final storage volume in period $t$, the probabilities of these same respective discrete storage volumes must also be equal. Thus,

$$\sum_j P_{l,j,t+1} = \sum_k \sum_i P_{kit} \quad \forall l, t \tag{8.41}$$

where the sums in the right hand side of Equation 8.41 are over only those combinations of $k$ and $i$ that result in a final volume interval $l$. This relationship is defined by Equation 8.39 ($l = \ell(k, i, t)$).

While Equation 8.41 must apply, it is not sufficient. The joint probability of a final storage volume in interval $l$ in period $t$ and an inflow $j$ in period $t + 1$ must equal the joint probability of an initial storage volume in the same interval $l$ and an inflow in the same interval $j$ in period $t + 1$. Multiplying the joint probability $P_{kit}$ times the conditional probability $P^t_{ij}$ and then summing over all $k$ and $i$ that results in a final storage interval $l$ defines the former, and the joint probability $P_{l,j,t+1}$ defines the latter.

$$P_{l,j,t+1} = \sum_k \sum_i P_{kit} P^t_{ij} \quad \forall l, j, t \quad l = \ell(k, i, t) \tag{8.42}$$

Once again the sums in Equation 8.42 are over all combinations of $k$ and $i$ that result in the designated storage volume interval $l$ as defined by the policy $\ell(k, i, t)$.

Finally, the sum of all joint probabilities $P_{kit}$ in each period $t$ must equal 1.

$$\sum_k \sum_i P_{kit} = 1 \quad \forall t \tag{8.43}$$

Note the similarity of Equations 8.42 and 8.43 to the Markov steady-state flow Equations 8.22 and 8.23. Instead of only one flow interval index considered in Equations 8.22 and 8.23, Equations 8.42 and 8.43 include two indices, one for storage volume intervals and the other for flow intervals. In both cases, one of Equations 8.22 and 8.42 can be omitted in each period $t$ since it is redundant with that period's Equations 8.23 and 8.43 respectively.

The unconditional probabilities $PS_{kt}$ and $PQ_{it}$ can be derived from the joint probabilities $P_{kit}$.

$$PS_{kt} = \sum_i P_{kit} \quad \forall k, t \tag{8.44}$$

$$PQ_{it} = \sum_k P_{kit} \quad \forall i, t \tag{8.45}$$

Each of these unconditional joint or marginal probabilities, when summed over all their volume and flow indices, will equal 1. For example,

$$\sum_k PS_{kt} = \sum_i PQ_{it} = 1 \tag{8.46}$$

Note that these probabilities are determined only on the basis of the relationships among flow and storage intervals as defined by Equation 8.39, $l = \ell(k, i, t)$ in each period $t$, and the Markov chains defining the flow interval transition or conditional probabilities, $P^t_{ij}$. It is not necessary to know the actual discrete storage values representing those intervals. Thus assuming any relationship among the storage volume and flow interval indices, $l = \ell(k, i, t)$ and a knowledge of the flow interval transition probabilities $P_{ij}^t$, one can determine the joint probabilities $P_{kit}$ and their marginal or unconditional probabilities $PS_{kt}$. One does not need to know what those storage intervals are to calculate their probabilities.

Given the values of these joint probabilities $P_{kit}$, the deterministic model defined by Equations 8.24 to 8.28 can be converted to a stochastic model to identify the best storage and allocation decision-variable values associated with each storage interval $k$ and flow interval $i$ in each period $t$.

$$\text{Minimize} \sum_k \sum_i \sum_{\tau}^{T} P_{kit}\{(UT_t - u_{kit})^2 + (DT_t - d_{kit})^2\} \tag{8.47}$$

The constraints include:

a) Continuity of storage involving initial storage volumes $S_{kt}$, net inflows $q_{it} - u_{kit}$, and at least partial releases $d_{kit}$. Again assuming no losses:

$$S_{kt} + q_{it} - u_{kit} - d_{kit} \geq S_{l,t+1} \quad \forall k, i, t$$
$$l = \ell(k, i, t) \tag{8.48}$$

b) Reservoir capacity limitations.

$$S_{kit} \leq K \quad \forall k, i, t \tag{8.49}$$

c) Allocation restrictions.

$$u_{kit} \leq q_{it} \quad \forall k, i, t \tag{8.50}$$

More detail on these and other stochastic modelling approaches can be found in Faber and Stedinger (2001); Gablinger and Loucks (1970); Huang et al. (1991); Kim and Palmer (1997); Loucks and Falkson (1970); Stedinger et al. (1984); Su and Deininger (1974); Tejada-Guibert et al. (1993 1995); and Yakowitz (1982).

## 6.2. A Numerical Example

A simple numerical example may help to illustrate how these stochastic models can be developed without getting buried in detail. Consider two within-year periods each year. The random flows $Q_t$ in each period $t$ are divided into two intervals. These flow intervals are represented by discrete flows of 1 and 3 volume units per second in the first period, and 3 and 6 volume units per second in the second period. Their transition probabilities are shown in Table 8.5.

| $Q_i$ flow in t = 1 | $Q_j$ flow in t = 2: 3 | 6 |
|---|---|---|
| 1 | 0.6 | 0.4 |
| 3 | 0.3 | 0.7 |

| $Q_i$ flow in t = 2 | $Q_j$ flow in t = 1: 1 | 3 |
|---|---|---|
| 3 | 0.7 | 0.3 |
| 6 | 0.2 | 0.8 |

**Table 8.5.** Transition probabilities for two ranges of flows in two within-year periods.

Assuming equal within-year period durations, these three discrete flow rates are equivalent to about 16, 47 and 95 million volume units per period.

Assume the active storage volume capacity $K$ in the reservoir equals 50 million volume units. This capacity can be divided into different intervals of storage. For this simple example, assume three storage volume intervals represented by 10, 25 and 40 million volume units. Assume the allocation targets remain the same in each period at both the upstream and downstream sites. The upstream allocation target is approximately 2 volume units per second or 30 million volume units in each period. The downstream allocation target is approximately 5 volume units per second or 80 million volume units in each period.

With these data we can use Equations 8.34 – 8.36 to determine the allocations that minimize the sum of squared deviations from targets and what that sum is, for all feasible combinations of initial and final storage volumes and flows. Table 8.6 shows the results of these optimizations. These results will be used in the dynamic programming model to determine the best final storage volumes given initial volumes and flows.

With the information in Tables 8.5 and 8.6, the dynamic programming model, Equation 8.38 or as expressed in Equation 8.51, can be solved to find the optimal final storage volumes, given an initial storage volume and flow. The iterations of the recursive equation, sufficient to reach a steady state, are shown in Table 8.7.

$$F_t^n(k, i) = \min_l \left\{ SD_{kil} + \sum_j P_{ij}^t F_{t+1}^{n-1}(l, j) \right\}$$

$$\text{such that} \quad S_{l,t+1} \leq K$$

$$S_{l,t+1} \leq S_{kt} + Q_{it} \tag{8.51}$$

E020829b

| initial storage $S_k$ | flow $Q_i$ | final storage $S_l$ | interval indices $k, i, l$ | upstream allocation $u_{ki}$ | down-stream allocation $d_{kil}$ | sum squared deviation $SD_{kil}$ |
|---|---|---|---|---|---|---|
| 10 | 16 | 10 | 1, 1, 1 | 0.0 | 16.0 | 4996.0 |
| 10 | 16 | 25 | 1, 1, 2 | 0.0 | 1.0 | 7141.0 |
| 10 | 47 | 10 | 1, 2, 1 | 0.0 | 47.0 | 1989.0 |
| 10 | 47 | 25 | 1, 2, 2 | 0.0 | 32.0 | 3204.0 |
| 10 | 47 | 40 | 1, 2, 3 | 0.0 | 17.0 | 4869.0 |
| 10 | 95 | 10 | 1, 3, 1 | 22.5 | 72.5 | 112.5 |
| 10 | 95 | 25 | 1, 3, 2 | 15.0 | 65.0 | 450.0 |
| 10 | 95 | 40 | 1, 3, 3 | 7.5 | 75.5 | 1012.5 |
| 25 | 16 | 10 | 2, 1, 1 | 0.0 | 31.0 | 3301.0 |
| 25 | 16 | 25 | 2, 1, 2 | 0.0 | 16.0 | 4996.0 |
| 25 | 16 | 40 | 2, 1, 3 | 0.0 | 1.0 | 7141.0 |
| 25 | 47 | 10 | 2, 2, 1 | 6.0 | 56.0 | 1152.0 |
| 25 | 47 | 25 | 2, 2, 2 | 0.0 | 47.0 | 1989.0 |
| 25 | 47 | 40 | 2, 2, 3 | 0.0 | 32.0 | 3204.0 |
| 25 | 95 | 10 | 2, 3, 1 | 30.0 | 80.0 | 0.0 |
| 25 | 95 | 25 | 2, 3, 2 | 22.5 | 72.5 | 112.5 |
| 25 | 95 | 40 | 2, 3, 3 | 15.0 | 65.0 | 450.0 |
| 40 | 16 | 10 | 3, 1, 1 | 0.0 | 46.0 | 2056.0 |
| 40 | 16 | 25 | 3, 1, 2 | 0.0 | 31.0 | 3301.0 |
| 40 | 16 | 40 | 3, 1, 3 | 0.0 | 16.0 | 4996.0 |
| 40 | 47 | 10 | 3, 2, 1 | 13.5 | 63.5 | 544.5 |
| 40 | 47 | 25 | 3, 2, 2 | 6.0 | 56.0 | 1152.0 |
| 40 | 47 | 40 | 3, 2, 3 | 0.0 | 47.0 | 1989.0 |
| 40 | 95 | 10 | 3, 3, 1 | 30.0 | 80.0 | 0.0 |
| 40 | 95 | 25 | 3, 3, 2 | 30.0 | 80.0 | 0.0 |
| 40 | 95 | 40 | 3, 3, 3 | 22.5 | 72.5 | 112.5 |

**Table 8.6.** Optimal allocations associated with given initial storage, $S_k$, flow, $Q_i$, and final storage, $S_l$, volumes. These allocations $u_{ki}$ and $d_{kil}$ minimize the sum of squared deviations, $DS_{kil} = (30 - u_{ki})^2 + (80 - d_{kil})^2$, from upstream and downstream targets, 30 and 80 respectively, subject to $u_{ki} \leq$ flow $Q_i$, and $d_{kil} \leq$ release $(S_k + Q_i - u_{ki} - S_l)$.

This process can continue until a steady-state policy is defined. Table 8.8 summarizes the next five iterations. At this stage, the annual differences in the objective values associated with a particular state and season have come close to a common constant value.

While the differences between corresponding $F_t^{n+T}$ and $F_t^n$ have not yet reached a common constant value to the nearest unit deviation (they range from, 3475.5 to 3497.1 for an average of 3485.7), the policy has converged to that shown in Tables 8.8 and 8.9.

Given this operating policy, the probabilities of being in any of these volume and flow intervals can be determined by solving Equations 8.42 through 8.45. Table 8.10 shows the results of these equations applied to the data in Tables 8.5 and 8.8. It is obvious that if the policy from Table 8.9 is followed, the steady-state probabilities of being in storage Interval 1 in Period 1 and in Interval 3 in Period 2 are 0.

Multiplying these joint probabilities by the corresponding $SD_{kit}$ values in the last column of Table 8.6 provides the annual expected squared deviations, associated with the selected discrete storage volumes and flows. This is done in Table 8.11 for those combinations of $k$, $i$, and $l$ that are contained in the optimal solution as listed in Table 8.9.

The sum of products of the last two columns in Table 8.11 for each period $t$ equals the expected squared deviations in the period. For period $t = 1$, the expected

**Table 8.7.** First four iterations of dynamic programming model, Equations 8.51, moving backward in successive periods $n$, beginning in season $t = 2$ with $n = 1$. The iterations stop when the final storage policy given any initial storage volume and flow repeats itself in two successive years. Initially, with no more periods remaining, $F_1^0(k, i) = 0$ for all $k$ and $i$.

E020829c

**period t = 2, n = 1**

| storage & flow $k, i$ | | $SD_{kil} + \Sigma_j P_{ij}^t F_{t+1}^{n-1}(l, j)$ | $F_t^n(k, i)$ | optimal $l$ |
|---|---|---|---|---|
| 1,2 | $l = 1$ | 1989.0 + 0 | 1989.0 | 1 |
| | $l = 2$ | 3204.0 + 0 | | |
| | $l = 3$ | 4869.0 + 0 | | |
| 1,3 | $l = 1$ | 112.5 + 0 | 112.5 | 1 |
| | $l = 2$ | 450.0 + 0 | | |
| | $l = 3$ | 1012.0 + 0 | | |
| 2,2 | $l = 1$ | 1152.0 + 0 | 1152.0 | 1 |
| | $l = 2$ | 1989.0 + 0 | | |
| | $l = 3$ | 3204.0 + 0 | | |
| 2,3 | $l = 1$ | 0.0 + 0 | 0.0 | 1 |
| | $l = 2$ | 112.5 + 0 | | |
| | $l = 3$ | 450.0 + 0 | | |
| 3,2 | $l = 1$ | 544.5 + 0 | 544.5 | 1 |
| | $l = 2$ | 1152.0 + 0 | | |
| | $l = 3$ | 1989.0 + 0 | | |
| 3,3 | $l = 1$ | 0.0 + 0 | 0.0 | 1,2 |
| | $l = 2$ | 0.0 + 0 | | |
| | $l = 3$ | 112.5 + 0 | | |

E020829d

**period t = 1, n = 2**

| storage & flow $k, i$ | | $SD_{kil} + \Sigma_j P_{ij}^t F_{t+1}^{n-1}(l, j)$ | $F_t^n(k, i)$ | optimal $l$ |
|---|---|---|---|---|
| 1,1 | $l = 1$ | 4996.0 + 0.6 (1989.0) +0.4 (112.5) = **6234.4** | 6234.4 | 1 |
| | $l = 2$ | 7141.0 + 0.6 (1152.0) +0.4 ( 0.0) = 7832.2 | | |
| | $l = 3$ | infeasible ---- = --- | | |
| 1,2 | $l = 1$ | 1989.0 + 0.3 (1989.0) +0.7 (112.5) = **2664.45** | 2664.5 | 1 |
| | $l = 2$ | 3204.0 + 0.3 (1152.0) +0.7 ( 0.0) = 3549.6 | | |
| | $l = 3$ | 4869.0 + 0.3 ( 544.5) +0.7 ( 0.0) = 5032.35 | | |
| 2,1 | $l = 1$ | 3301.0 + 0.6 (1989.0) +0.4 (112.5) = **4539.4** | 4539.4 | 1 |
| | $l = 2$ | 4996.0 + 0.6 (1152.0) +0.4 ( 0.0) = 5687.2 | | |
| | $l = 3$ | 7141.0 + 0.6 ( 544.5) +0.4 ( 0.0) = 7467.7 | | |
| 2,2 | $l = 1$ | 1152.0 + 0.3 (1989.0) +0.7 (112.5) = **1827.45** | 1827.5 | 1 |
| | $l = 2$ | 1989.0 + 0.3 (1152.0) +0.7 ( 0.0) = 2334.6 | | |
| | $l = 3$ | 3204.0 + 0.3 ( 544.5) +0.7 ( 0.0) = 3367.35 | | |
| 3,1 | $l = 1$ | 2056.0 + 0.6 (1989.0) +0.4 (112.5) = **3294.4** | 3294.4 | 1 |
| | $l = 2$ | 3301.0 + 0.6 (1152.0) +0.4 ( 0.0) = 3992.2 | | |
| | $l = 3$ | 4996.0 + 0.6 ( 544.5) +0.4 ( 0.0) = 5322.7 | | |
| 3,2 | $l = 1$ | 544.5 + 0.3 (1989.0) +0.7 (112.5) = **1219.95** | 1220.0 | 1 |
| | $l = 2$ | 1152.0 + 0.3 (1152.0) +0.7 ( 0.0) = 1497.6 | | |
| | $l = 3$ | 1989.0 + 0.3 ( 544.5) +0.7 ( 0.0) = 2152.35 | | |

*(contd.)*

**Table 8.7.** Concluded.

E020829e

| storage & flow $k, i$ | | period $t = 2$, $n = 3$ — $SD_{kil} + \Sigma_j P_{ij}^t F_{t+1}^{n-1}(l, j)$ | $F_t^n(k, i)$ | optimal $l$ |
|---|---|---|---|---|
| 1,2 | $l = 1$ | 1989.0 + 0.7 (6234.4) +0.3 (2664.5) = 7152.4 | | |
| | $l = 2$ | 3204.0 + 0.7 (4539.4) +0.3 (1827.5) = **6929.8** | 6929.8 | 2 |
| | $l = 3$ | 4869.0 + 0.7 (3294.4) +0.3 (1219.9) = 7541.1 | | |
| 1,3 | $l = 1$ | 112.5 + 0.2 (6234.4) +0.8 (2664.5) = 3490.0 | | |
| | $l = 2$ | 450.0 + 0.2 (4539.4) +0.8 (1827.5) = 2819.8 | | |
| | $l = 3$ | 1012.5 + 0.2 (3294.4) +0.8 (1219.9) = **2647.3** | 2647.3 | 3 |
| 2,2 | $l = 1$ | 1152.0 + 0.7 (6234.4) +0.3 (2664.5) = 6315.4 | | |
| | $l = 2$ | 1989.0 + 0.7 (4539.4) +0.3 (1827.5) = **5714.8** | 5714.8 | 2 |
| | $l = 3$ | 3204.0 + 0.7 (3294.4) +0.3 (1219.9) = 5876.1 | | |
| 2,3 | $l = 1$ | 0.0 + 0.2 (6234.4) +0.8 (2664.5) = 3378.4 | | |
| | $l = 2$ | 112.5 + 0.2 (4539.4) +0.8 (1827.5) = 2482.3 | | |
| | $l = 3$ | 450.0 + 0.2 (3294.4) +0.8 (1219.9) = **2084.8** | 2084.8 | 3 |
| 3,2 | $l = 1$ | 544.5 + 0.7 (6234.4) +0.3 (2664.5) = 5707.9 | | |
| | $l = 2$ | 1152.0 + 0.7 (4539.4) +0.3 (1827.5) = 4877.8 | | |
| | $l = 3$ | 1989.0 + 0.7 (3294.4) +0.3 (1219.9) = **4661.1** | 4661.1 | 3 |
| 3,3 | $l = 1$ | 0.0 + 0.2 (6234.4) +0.8 (2664.5) = 3378.4 | | |
| | $l = 2$ | 0.0 + 0.2 (4539.4) +0.8 (1827.5) = 2369.8 | | |
| | $l = 3$ | 112.5 + 0.2 (3294.4) +0.8 (1219.9) = **1747.3** | 1747.3 | 3 |

E020829f

| storage & flow $k, i$ | | period $t = 1$, $n = 4$ — $SD_{kil} + \Sigma_j P_{ij}^t F_{t+1}^{n-1}(l, j)$ | $F_t^n(k, i)$ | optimal $l$ |
|---|---|---|---|---|
| 1,1 | $l = 1$ | 4996.0 + 0.6 (6929.8) +0.4 (2647.3) = **10212.8** | 10212.8 | 1 |
| | $l = 2$ | 7141.0 + 0.6 (5714.8) +0.4 (2084.8) = 11403.8 | | |
| | $l = 3$ | infeasible --- | | |
| 1,2 | $l = 1$ | 1989.0 + 0.3 (6929.8) +0.7 (2647.3) = **5921.1** | 5921.1 | 1 |
| | $l = 2$ | 3204.0 + 0.3 (5714.8) +0.7 (2084.8) = 6377.8 | | |
| | $l = 3$ | 4869.0 + 0.3 (4661.1) +0.7 (1747.3) = 7490.5 | | |
| 2,1 | $l = 1$ | 3301.0 + 0.6 (6929.8) +0.4 (2647.3) = **8517.8** | 8517.8 | 1 |
| | $l = 2$ | 4996.0 + 0.6 (5714.8) +0.4 (2084.8) = 9258.8 | | |
| | $l = 3$ | 7141.0 + 0.6 (4661.1) +0.4 (1747.3) = 10636.6 | | |
| 2,2 | $l = 1$ | 1152.0 + 0.3 (6929.8) +0.7 (2647.3) = **5084.1** | 5084.1 | 1 |
| | $l = 2$ | 1989.0 + 0.3 (5714.8) +0.7 (2084.8) = 5162.8 | | |
| | $l = 3$ | 3204.0 + 0.3 (4661.1) +0.7 (1747.3) = 5825.5 | | |
| 3,1 | $l = 1$ | 2056.0 + 0.6 (6929.8) +0.4 (2647.3) = **7272.8** | 7272.8 | 1 |
| | $l = 2$ | 3301.0 + 0.6 (5714.8) +0.4 (2084.8) = 7563.8 | | |
| | $l = 3$ | 4996.0 + 0.6 (4661.1) +0.4 (1747.3) = 8491.6 | | |
| 2,2 | $l = 1$ | 544.5 + 0.3 (6929.8) +0.7 (2647.3) = 4476.6 | | |
| | $l = 2$ | 1152.0 + 0.3 (5714.8) +0.7 (2084.8) = **4325.8** | 4325.8 | 2 |
| | $l = 3$ | 1989.0 + 0.3 (4661.1) +0.7 (1747.3) = 4610.5 | | |

**Table 8.8.** Summary of objective function values $F_t^n(k, i)$ and optimal decisions for stages $n = 5$ to 9 periods remaining.

| storage & flow $k, i$ | $t = 2, n = 5$ $F_t^n(k,i)$ | $l^*$ | $t = 1, n = 6$ $F_t^n(k,i)$ | $l^*$ | $t = 2, n = 7$ $F_t^n(k,i)$ | $l^*$ | $t = 1, n = 8$ $F_t^n(k,i)$ | $l^*$ | $t = 2, n = 9$ $F_t^n(k,i)$ | $l^*$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1,1 | | | 13782.1 | 1 | | | 17279.2 | 1 | | |
| 1,2 | 10691.7 | 2 | 9345.9 | 1 | 14217.7 | 2 | 12821.4 | 1 | 17708.3 | 2 |
| 1,3 | 5927.7 | 3 | | | 9381.5 | 3 | | | 12861.3 | 3 |
| 2,1 | | | 12087.1 | 1 | | | 15584.2 | 1 | | |
| 2,2 | 9476.7 | 2 | 8508.9 | 1 | 13002.7 | 2 | 11984.4 | 1 | 16493.2 | 2 |
| 2,3 | 5365.2 | 3 | | | 8819.0 | 3 | | | 12298.7 | 3 |
| 3,1 | | | 10842.1 | 1 | | | 14339.2 | 1 | | |
| 3,2 | 8377.7 | 3 | 7750.7 | 2 | 11903.7 | 3 | 11226.1 | 2 | 15394.3 | 3 |
| 3,3 | 5027.7 | 3 | | | 8481.5 | 3 | | | 11961.2 | 3 |

E020829g

sum of squared deviations are 1893.3 and for $t = 2$ they are 1591.0. The total annual expected squared deviations are 3484.3. This compares with the expected squared deviations derived from the dynamic programming model, after 9 iterations, ranging from 3475.5 to 3497.1 (as calculated from data in Table 8.8).

These upstream allocation policies can be displayed in plots, as shown in Figure 8.12.

The policy for reservoir releases is a function not only of the initial storage volumes, but also of the current inflow, in other words, the total water available in the period. Reservoir release rule curves such as shown in Figures 4.16 or 4.18 now must become two-dimensional. However, the inflow for each period usually cannot be predicted with certainty at the beginning of each period. In situations where the release cannot be adjusted during the period as the inflow becomes more predictable, the reservoir release policy has to be expressed in a way that can be followed without knowledge of the current inflow. One way to do this is to compute the expected value of the release for each discrete storage volume, and show it in a release rule. This is done in Figure 8.13. The probability of each discrete release associated with each discrete river flow is the probability of the flow itself. Thus, in Period 1 when the storage volume is 40, the expected release is 46(0.41) + 56(0.59) = 52. These discrete expected releases can be used to define a continuous range of releases for the continuous range of storage volumes from 0 to full capacity, 50. Figure 8.13 also shows the hedging that might take place as the reservoir storage volume decreases.

Another approach to defining the releases in each period in a manner that is not dependent on knowledge of the current inflow, even though the model used assumes this, is to attempt to define either release targets with constraints on final storage volumes, or final storage targets with constraints on total releases. Obviously, such policies will not guarantee constant releases throughout each period. For example, consider the optimal policy shown in Table 8.9. The releases (or final storage volumes) in each period are dependent on the initial storage and current inflow. However, this operating policy can be expressed as:

- If in period 1, the final storage target should be in interval 1. Yet the total release cannot exceed the flow in interval 2.
- If in period 2 and the initial storage is in interval 1, the release should be in interval 1.
- If in period 2 and the initial storage is in interval 2, the release should be in interval 2.
- If in period 2 and the initial storage is in interval 3, the release should equal the inflow.

This policy can be followed without any forecast of current inflow. It will provide the releases and final storage volumes that would be obtained with a perfect inflow forecast at the beginning of each period.

Table 8.9. Optimal reservoir policy $l = \ell(k, i, t)$ for the example problem.

| period $t$ | initial storage volume and flow interval $k$ | $i$ | final storage volume interval $l$ |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 1 | 1 | 2 | 1 |
| 1 | 2 | 1 | 1 |
| 1 | 2 | 2 | 1 |
| 1 | 3 | 1 | 1 |
| 1 | 3 | 2 | 2 |
| 2 | 1 | 2 | 2 |
| 2 | 1 | 3 | 3 |
| 2 | 2 | 2 | 2 |
| 2 | 2 | 3 | 3 |
| 2 | 3 | 2 | 3 |
| 2 | 3 | 3 | 3 |

E020829h

Table 8.10. Probabilities of flow and storage volume intervals associated with the policy as defined in Table 8.9 for the example problem.

E020829j

**unconditional probabilities $PQ_{it}$ of flow intervals $i$ in the 2 time periods $t$**

| | |
|---|---|
| $PQ(1, 1) = 0.4117647$ | $PQ(2, 2) = 0.4235294$ |
| $PQ(2, 1) = 0.5882353$ | $PQ(3, 2) = 0.5764706$ |

**unconditional probabilities $PS_{kt}$ of storage intervals $k$ in the 2 time periods $t$**

| | |
|---|---|
| $PS(1, 1) = 0.0000000$ | $PS(1, 2) = 0.5388235$ |
| $PS(2, 1) = 0.4235294$ | $PS(2, 2) = 0.4611765$ |
| $PS(3, 1) = 0.5764706$ | $PS(3, 2) = 0.0000000$ |

**joint probabilities $P_{kit}$ of storage volume intervals $k$ and flow intervals $i$ in the 2 time periods $t$**

| | |
|---|---|
| $P(1, 1, 1) = 0.0000000$ | $P(1, 2, 2) = 0.2851765$ |
| $P(1, 2, 1) = 0.0000000$ | $P(1, 3, 2) = 0.2536471$ |
| $P(2, 1, 1) = 0.2964706$ | $P(2, 2, 2) = 0.1383529$ |
| $P(2, 2, 1) = 0.1270588$ | $P(2, 3, 2) = 0.3228235$ |
| $P(3, 1, 1) = 0.1152941$ | $P(3, 2, 2) = 0.0000000$ |
| $P(3, 2, 1) = 0.4611765$ | $P(3, 3, 2) = 0.0000000$ |

**Table 8.11.** The optimal operating policy and the probability of each state and decision.

| initial storage $S_k$ | flow $Q_i$ | final storage $S_l$ | interval indices $k, i, l$ | time period | optimal allocation $u_{kit}$ | optimal allocation $d_{kit}$ | sum squared deviations $SD_{kit}$ | joint probability $P_{kit}$ |
|---|---|---|---|---|---|---|---|---|
| 10 | 16 | 10 | 1,1,1 | 1 | 0.0 | 16.0 | 4996.0 | 0.0 |
| 10 | 47 | 10 | 1,2,1 | 1 | 0.0 | 47.0 | 1989.0 | 0.0 |
| 25 | 16 | 10 | 2,1,1 | 1 | 0.0 | 31.0 | 3301.0 | 0.2964706 |
| 25 | 47 | 10 | 2,2,1 | 1 | 6.0 | 56.0 | 1152.0 | 0.1270588 |
| 40 | 16 | 10 | 3,1,1 | 1 | 0.0 | 46.0 | 2056.0 | 0.1152941 |
| 40 | 47 | 25 | 3,2,2 | 1 | 6.0 | 56.0 | 1152.0 | 0.4611765 |
| | | | | | | | sum = | 1.0 |
| 10 | 47 | 25 | 1,2,2 | 2 | 0.0 | 32.0 | 3204.0 | 0.2851765 |
| 10 | 95 | 25 | 1,3,2 | 2 | 7.5 | 57.5 | 1012.5 | 0.2536471 |
| 25 | 47 | 25 | 2,2,2 | 2 | 0.0 | 47.0 | 1989.0 | 0.1383529 |
| 25 | 95 | 40 | 32,3 | 2 | 15.0 | 65.0 | 450.0 | 0.3228235 |
| 40 | 47 | 40 | 3,2,3 | 2 | 0.0 | 47.0 | 1989.0 | 0.0 |
| 40 | 95 | 40 | 3,3,3 | 2 | 22.5 | 72.5 | 112.5 | 0.0 |
| | | | | | | | sum = | 1.0 |

E020829k

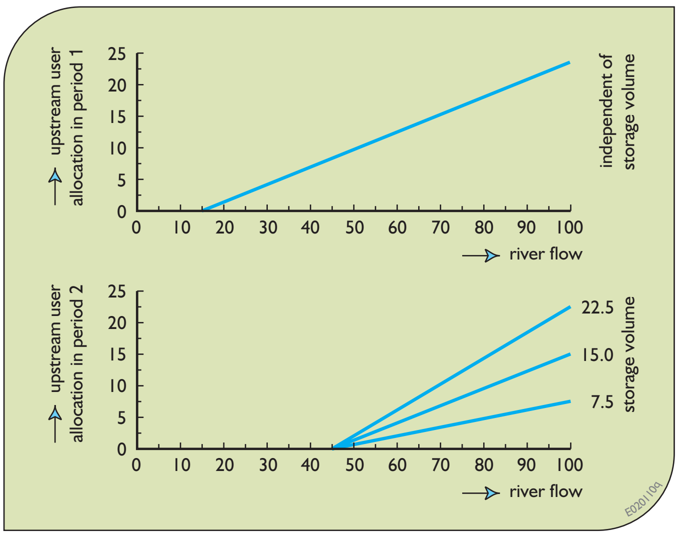



**Figure 8.12.** Upstream user allocation policies. In Period 1 they are independent of the downstream initial storage volumes. In Period 2 the operator would interpolate between the three allocation functions given for the three discrete initial reservoir storage volumes.

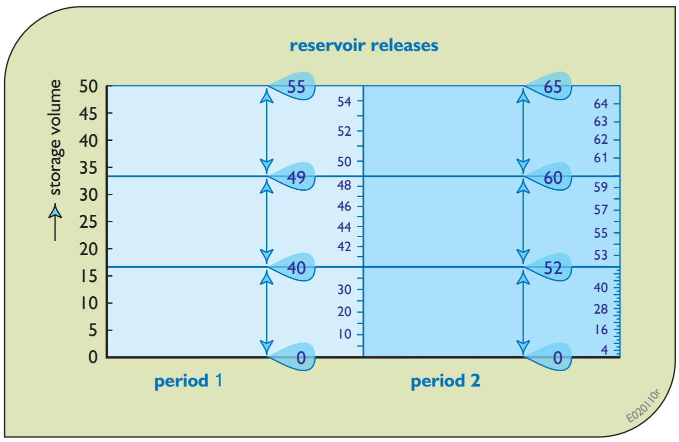



**Figure 8.13.** Reservoir release rule showing an interpolated release, increasing as storage volumes increase.

Alternatively, in each period $t$ one can solve the model defined by Equation 8.37 to obtain the best decision for the current and a sequence of future periods, taking into account all current information regarding the objectives and possible inflow scenarios and their probabilities. The actual release decision in the current period can be the expected value of all these releases in this current period. At the beginning of the next period, the model is updated with respect to current initial storage and inflow scenarios (as well as any changes in objectives or other constraints) and solved again. This process continues in real time. This approach is discussed in Tejada-Guibert et al. (1993) and is the current approach for providing release advice to the Board of Control that oversees the releases from Lake Ontario that govern the water levels of the lake and the St. Lawrence River.

These policies, and modifications of them, can be simulated to determine improved release rules.

## 7. Conclusions

This chapter has introduced some approaches for including risk in optimization and simulation models. The discussion began with ways to obtain values of random variables whose probability distributions are known. These values, for example streamflows or parameter values, can be inputs to simulation models. Monte Carlo simulation involves the use of multiple simulations using these random variable values to obtain the probability distributions of outputs, including various system performance indicators.

Two methods were reviewed for introducing random variables along with their probabilities into optimization models. One involves the use of chance constraints. These are constraints that must be met, as all constraints must, but now with a certain probability. As in any method there are limits to the use of chance constraints. These limitations were not discussed, but in cases where chance constraints are applicable, and if their deterministic equivalents can be defined, they are probably the only method of introducing risk into otherwise deterministic models that do not add to the model size.

Alternatively, the range of random variable values can be divided into discrete ranges. Each range can be represented by a specific or discrete value of the random variable. These discrete values and their probabilities can become part of an optimization model. This was demonstrated by means of transition probabilities incorporated into both linear and dynamic programming models.

The examples used in this chapter to illustrate the development and application of stochastic optimization and simulation models are relatively simple. These and similar probabilistic and stochastic models have been applied to numerous water resources planning and management problems. They can be a much more effective screening tool than deterministic models based on the mean or other selected values of random variables. But sometimes they are not. Clearly if the system being analysed is very complex, or just very big in terms of the number of variables and constraints, the use of deterministic models for a preliminary screening of alternatives prior to a more precise probabilistic screening is often warranted.

## 8. References

FABER, B.A. and STEDINGER, J.R. 2001. Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts. *Journal of Hydrology*, Vol. 249, Nos. 1–4, pp. 113–33.

GABLINGER, M. and LOUCKS, D.P. 1970. Markov models for flow regulation. *Journal of the Hydraulics Division, ASCE,* Vol. 96, No. HY1, pp. 165–81.

HUANG, W.C.; HARBO, R. and BOGARDI, J.J. 1991. Testing stochastic dynamic programming models conditioned on observed or forecast inflows. *Journal of Water Resources Planning and Management*, Vol. 117, No.1, pp. 28–36.

KIM, Y.O.; PALMER, R.N. 1997. Value of seasonal flow forecasts in bayesian stochastic programming. *Journal of Water Resources Planning and Management*, Vol. 123, No. 6, pp. 327–35.

LOUCKS, D.P. and FALKSON, L.M. 1970. Comparison of some dynamic, linear, and policy iteration methods for reservoir operation. *Water Resources Bulletin*, Vol. 6, No. 3, pp. 384–400.

STEDINGER, J.R.; SULE, B.F. and LOUCKS, D.P. 1984. Stochastic dynamic programming models for reservoir operation optimization. *Water Resources Research*, Vol. 20, No. 11, pp. 1499–505.

SU, S.Y. and DEININGER, R.A. 1974. Modeling the regulation of lake superior under uncertainty of future water supplies. *Water Resources Research*, Vol. 10, No. 1, pp. 11–22.

TEJADA-GUIBERT, J.A.; JOHNSON, S.A. and STEDINGER, J.R. 1993. Comparison of two approaches for implementing multi-reservoir operating policies derived using dynamic programming. *Water Resources Research*, Vol. 29, No. 12, pp. 3969–80.

TEJADA-GUIBERT, J.A.; JOHNSON, S.A. and STEDINGER, J.R. 1995. The value of hydrologic information in stochastic dynamic programming models of a multi-reservoir system. *Water Resources Research*, Vol. 31, No. 10, pp. 2571–9.

YAKOWITZ, S. 1982. Dynamic programming applications in water resources. *Water Resources Research*, Vol. 18, No. 4, pp. 673–96.

## Additional References (Further Reading)

BASSSON, M.S.; ALLEN, R.B.; PEGRAM, G.G.S. and VAN ROOYEN, J.A. 1994. *Probabilistic management of water resource and hydropower systems*. Highlands Ranch, Colo., Water Resources Publications.

BIRGE, J.R. and LOUVEAUX, F. 1997. *Introduction to stochastic programming*. New York, Springer-Verlag.

DEGROOT, M.H. 1980. *Optimal statistical decisions*. New York, McGraw-Hill.

DEMPSTER, M.A.H. (ed.). 1980. *Stochastic programming*. New York, Academic Press.

DORFMAN, R.; JACOBY, H.D. and THOMAS, H.A. Jr. 1972. *Models for managing regional water quality*. Cambridge, Mass., Harvard University Press.

ERMOLIEV, Y. 1970. *Methods of stochastic programming*. Moscow, Nauka. (In Russian.)

ERMOLIEV, Y. and WETS, R., (eds.). 1988. *Numerical techniques for stochastic optimization*. Berlin Springer-Verlag.

HEYMAN, D.P. and SOBEL, M.J. 1984. *Stochastic models in operations research. Vol. 2: Stochastic optimization*. New York, McGraw-Hill.

HILLIER, F.S. and LIEBERMAN, G.J. 1990. *Introduction to stochastic models in operations research*. New York, McGraw-Hill.

HOWARD, R.A. 1960. *Dynamic programming and Markov processes*. Cambridge, Mass., MIT Press.

HUFSCHMIDT, M.M. and FIERING, M.B. 1966. *Simulation techniques for design of water-resource systems*. Cambridge, Mass., Harvard University Press.

LOUCKS, D.P.; STEDINGER, J.S. and HAITH, D.A. 1981. *Water resource systems planning and analysis*. Englewood Cliffs, N.J., Prentice Hall.

MAASS, A.; HUFSCHMIDT, M.M.; DORFMAN, R.; THOMAS, H.A. Jr.; MARGLIN, S.A. and FAIR, G.M. 1962. *Design of water-resource systems*. Cambridge, Mass., Harvard University Press.

PRÉKOPA, A. 1995. *Stochastic programming*. Dordrecht, Kluwer Academic.

RAIFFA, H. and SCHLAIFER, R. 1961. *Applied statistical decision theory*. Cambridge, Mass., Harvard University Press.

REVELLE, C. 1999. *Optimizing reservoir resources*. New York, Wiley.

ROSS, S.M. 1983. *Introduction to stochastic dynamic programming*. New York, Academic Press.

SOMLYÓDY, L. and WETS, R.J.B. 1988. Stochastic optimization models for lake eutrophication management. *Operations Research*, Vol. 36, No. 5, pp. 660–81.

WETS, R.J.B. 1990 Stochastic programming. In: G.L. Nemhauser, A.H.G. Rinnoy Kan and M.J. Todd (eds.), *Optimization: handbooks in operations research and management science*, Vol. 1, pp. 573–629. Amsterdam, North-Holland.

WURBS, R.A. 1996. *Modeling and analysis of reservoir system operations*. Upper Saddle River, N.J., Prentice Hall.

# 9. Model Sensitivity and Uncertainty Analysis

1. Introduction 255

2. Issues, Concerns and Terminology 256

3. Variability and Uncertainty In Model Output 258
   3.1. Natural Variability 259
   3.2. Knowledge Uncertainty 260
      3.2.1. Parameter Value Uncertainty 260
      3.2.2. Model Structural and Computational Errors 260
   3.3. Decision Uncertainty 260

4. Sensitivity and Uncertainty Analyses 261
   4.1. Uncertainty Analyses 261
      4.1.1. Model and Model Parameter Uncertainties 262
      4.1.2. What Uncertainty Analysis Can Provide 265
   4.2. Sensitivity Analyses 265
      4.2.1. Sensitivity Coefficients 267
      4.2.2. A Simple Deterministic Sensitivity Analysis Procedure 267
      4.2.3. Multiple Errors and Interactions 269
      4.2.4. First-Order Sensitivity Analysis 270
      4.2.5. Fractional Factorial Design Method 272
      4.2.6. Monte Carlo Sampling Methods 273

5. Performance Indicator Uncertainties 278
   5.1. Performance Measure Target Uncertainty 278
   5.2. Distinguishing Differences Between Performance Indicator Distributions 281

6. Communicating Model Output Uncertainty 283

7. Conclusions 285

8. References 287

# 9 Model Sensitivity and Uncertainty Analysis

The usefulness of any model depends in part on the accuracy and reliability of its output. Yet, because all models are imperfect abstractions of reality, and because precise input data are rarely if ever available, all output values are subject to imprecision. Input data errors and modelling uncertainties are not independent of each other – they can interact in various ways. The end result is imprecision and uncertainty associated with model output. This chapter focuses on ways of identifying, quantifying, and communicating the uncertainties in model outputs.

## 1. Introduction

Models are the primary way we have to estimate the multiple effects of alternative water resources system design and operating policies. Models predict the values of various system performance indicators. Their outputs are based on model structure, hydrological and other time-series inputs, and a host of parameters whose values describe the system being simulated. Even if these assumptions and input data reflect, or are at least representative of, conditions believed to be true, we know they will be inaccurate. Our models are always simplifications of the real systems we study. Furthermore, we simply cannot forecast the future with precision, so we know the model outputs of future conditions are at best uncertain.

Some prediction uncertainties can be reduced by additional research and data collection and analysis. Before undertaking expensive studies to gather and analyse additional data, it is reasonable to ask what improvement in estimates of system performance, or what reduction in the uncertainty associated with those estimates, would result if all data and model uncertainties could be reduced. Such information helps determine how much one would be willing to 'pay' in order to reduce prediction uncertainty. If prediction uncertainty is costly, it may pay to invest in additional data collection, more studies or better models – all aimed at reducing that prediction uncertainty. If that uncertainty has no, or only a very modest, impact on the likely decision that is to be made, one should find other issues to worry about.

If it appears that reducing prediction uncertainty is worthwhile, then one should consider how best to do it. If it involves obtaining additional information, then it is clear that the value of this additional information, however measured, should exceed the cost of obtaining it. The value of such information will be the increase in system performance, or the reduction in its variance, that one can expect from obtaining such information. If additional information is to be obtained, it should be information that reduces the uncertainties considered important, not the unimportant ones.

This chapter reviews some methods for identifying and communicating model prediction uncertainty. The discussion begins with a review of the causes of risk and uncertainty in model output. It then examines ways of measuring or quantifying uncertainty and model output sensitivity to model input imprecision, concentrating on methods that seem most relevant or practical for large-scale regional simulation modelling. It builds on some of the statistical methods reviewed in Chapter 7 and the modelling of risk and uncertainty in Chapter 8.

## 2. Issues, Concerns and Terminology

Outcomes or events that cannot be predicted with certainty are often called risky or uncertain. Some individuals draw a special and interesting distinction between risk and uncertainty. In particular, the term *risk* is often reserved to describe situations for which probabilities are available to describe the likelihood of various events or outcomes. If probabilities of various events or outcomes cannot be quantified, or if the events themselves are unpredictable, some would say the problem is then one of *uncertainty*, and not of risk. In this chapter, what is not certain is considered uncertain, and uncertainty is often described by a probability distribution. When the ranges of possible events are known and their probabilities are measurable, risk is called *objective risk*. If the probabilities are based solely on human judgement, the risk is called *subjective risk*.

Such distinctions between objective and subjective risk, and between risk and uncertainty, rarely serve any useful purpose to those developing and using models. Likewise, the distinctions are often unimportant to those who should be aware of the risks or uncertainties associated with system performance indicator values.

Uncertainty in information is inherent in future-oriented planning efforts. It stems from inadequate information and incorrect assumptions, as well as from the variability of natural processes. Water managers often need to identify both the uncertainty and the sensitivity of, or changes in, system performance indicator values due to any changes in possible input data and parameter values from what were predicted. They need to reduce this level of uncertainty to the extent practicable. Finally, they need to communicate the residual uncertainties clearly so that decisions can be made with this knowledge and understanding.

Sensitivity analysis can be distinguished from uncertainty analysis. Sensitivity analysis procedures explore and quantify the impact of possible errors in input data on predicted model outputs and system performance indices. Simple sensitivity analysis procedures can be used to illustrate either graphically or numerically the consequences of alternative assumptions about the future. Uncertainty analyses employing probabilistic descriptions of model inputs can be used to derive probability distributions of model outputs and system performance indices. Figure 9.1 illustrates the impact of both input data sensitivity and input data uncertainty on model output uncertainty.

It is worthwhile to explore the transformation of uncertainties in model inputs and parameters into uncertainty in model outputs when conditions differ from those reflected by the model inputs. Historical records of system characteristics are typically used as a basis for model inputs. Yet conditions in the future may change. There

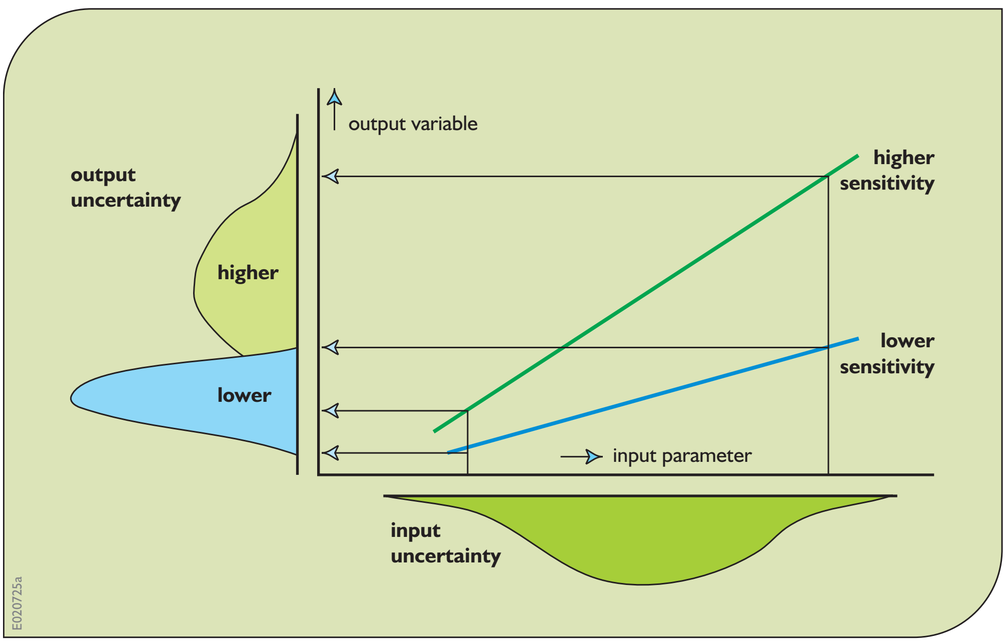



**Figure 9.1.** Schematic diagram showing relationship among model input parameter uncertainty and sensitivity to model output variable uncertainty (Lal, 1995).

may be changes in the frequency and amounts of precipitation, changes in land cover and topography, and changes in the design and operation of control structures, all resulting in changes of water stages and flows, and their qualities, and consequently changes in the affected ecosystems.

If asked how the system would operate with inputs similar to those in the historical database, the model should be able to interpolate within the available knowledge base to provide a fairly precise estimate. Still, that estimate will not be perfect. This is because our ability to reproduce current and recent operations is not perfect, though it should be fairly good. If asked to predict system performance for situations very different from those in the historical knowledge base, or when the historical data are not considered representative of what might happen in the future (due to climate change for example), such predictions become much less precise.

There are two reasons for this. First, our description of the characteristics of those different situations or conditions may be imprecise. Second, our knowledge base may not be sufficient to calibrate model parameters in ways that would enable us to reliably predict how the system will operate under conditions unlike those that have been experienced historically. The more conditions of interest differ from those in the historical knowledge base, the less confidence we have that the model is providing a reliable description of systems operation. Figure 9.2 illustrates this issue.

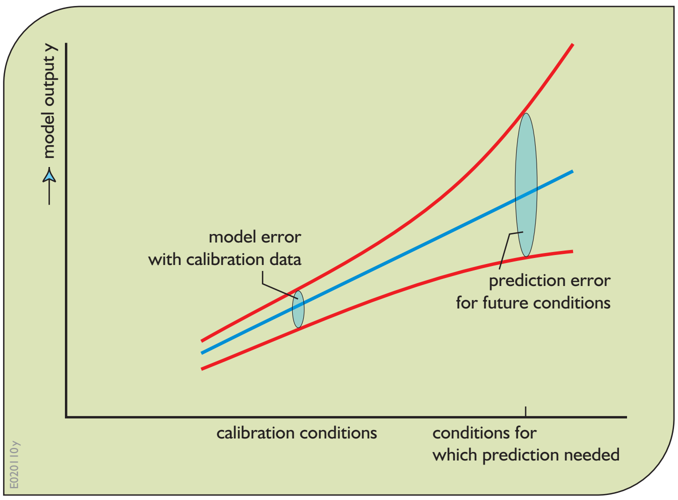



**Figure 9.2.** The precision of model predictions is affected by the difference between the conditions or scenarios of interest and the conditions or scenarios for which the model was calibrated.

Clearly, an uncertainty analysis needs to consider how well a model can replicate current operations, and how similar the target conditions or scenarios are to those described in the historical record. The greater the required extrapolation from what has been observed, the greater will be the importance of parameter and model uncertainties.

The relative and absolute importance of different parameters will depend on the system performance indicators of interest. Seepage rates may have a very large local impact, but a small global effect. Changes in system-wide evapotranspiration rates are likely to affect system-wide flows. The precision of model projections and the relative importance of errors in different parameters will depend upon:

- the precision with which the model can reproduce observed conditions
- the difference between the predicted conditions and the historical experience included in the knowledge base
- system performance characteristics of interest.

Errors and approximations in input data measurement, parameter values, model structure and model solution algorithms are all sources of uncertainty. While there are reasonable ways of quantifying and reducing these errors and the resulting range of uncertainty of various system performance indicator values, they are impossible to eliminate. Decisions will still have to be made in the face of a risky and uncertain future, and can be modified as new data and knowledge are obtained in a process of adaptive management.

There is also uncertainty with respect to human behaviour and reactions related to particular outcomes and their likelihoods, that is, to their risks and uncertainties. As important as risks and uncertainties associated with human reactions are to particular outcomes, they are not usually part of the models themselves. Social uncertainty may often be the most significant component of the total uncertainty associated with just how a water resources system will perform. For this reason we should seek designs and operating policies that are flexible and adaptable.

When uncertainties associated with system operation under a new operating regime are large, one should anticipate the need to make changes and improvements as experience is gained and new information accumulates. When predictions are highly unreliable, responsible managers should favour actions that are robust (good under a wide range of situations), gain information through research and experimentation, monitor results to provide feedback for the next decision, update assessments and modify policies in the light of new information, and avoid irreversible actions and commitments.

## 3. Variability and Uncertainty in Model Output

Differences between model output and observed values can result from either natural variability, such as is caused by unpredictable rainfall, evapotranspiration, water consumption and the like, and/or by both known and unknown errors in the input data, the model parameters or the model itself. The later is sometimes called knowledge uncertainty, but it is not always due to a lack of knowledge. Models are always simplifications of reality and, hence, 'imprecision' can result. Sometimes imprecision occurs because of a lack of knowledge, such as just how a particular species will react to various environmental and other habitat conditions. At other times, known errors are introduced simply for practical reasons.

Imperfect representation of processes in a model constitutes model *structural uncertainty*. Imperfect knowledge of the values of parameters associated with these processes constitutes model *parameter uncertainty*. *Natural variability* includes both *temporal variability* and *spatial variability*, to which model input values may be subject.

Figure 9.3 illustrates these different types of uncertainty. For example, the rainfall measured at a weather station within a particular model grid cell may be used as an input value for that cell, but the rainfall may actually vary at different points within that cell and its mean value will vary across the landscape. Knowledge uncertainty can be reduced through further measurement and/or research. Natural variability is a property of the natural system, and is usually not reducible at the scale being used. Decision uncertainty is simply an acknowledgement that we cannot predict ahead of time just what decisions individuals and organizations will make, or even just what particular set of goals or objectives will be considered and the relative importance of each.

Rather than contrasting 'knowledge' uncertainty versus natural variability versus decision uncertainty, one can classify uncertainty in another way based on specific sources of uncertainty, such as those listed below, and address ways of identifying and dealing with each source of uncertainty.

- Informational uncertainties
  – imprecision in specifying the boundary and initial conditions that impact the output variable values
  – imprecision in measuring observed output variable values.
- Model uncertainties
  – uncertain model structure and parameter values
  – variability of observed input and output values over a region smaller than the spatial scale of the model

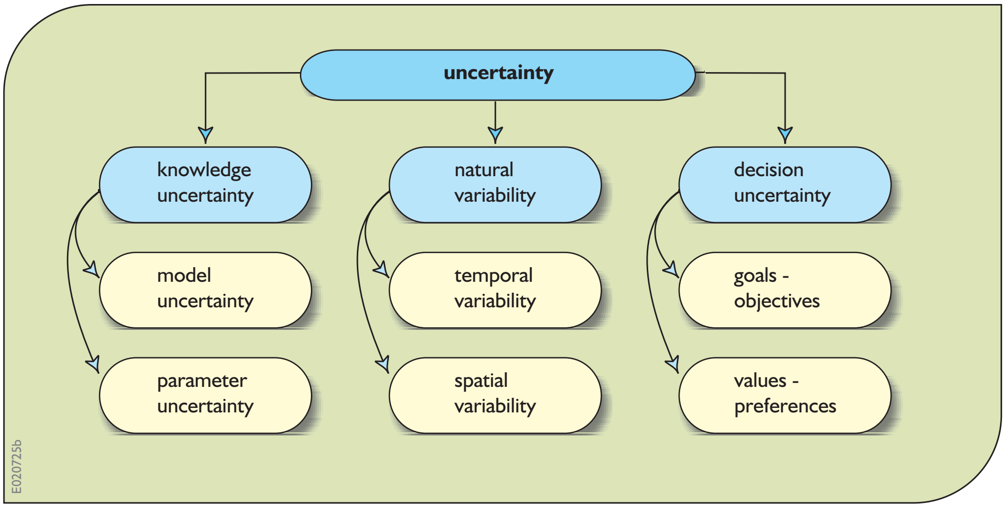



**Figure 9.3.** One way of classifying types of uncertainty.

  - variability of observed model input and output values within a time smaller than the temporal scale of the model. (e.g., rainfall and depths and flows within a day)
  - errors in linking models of different spatial and temporal scales.
- Numerical errors
  - errors in the model solution algorithm.

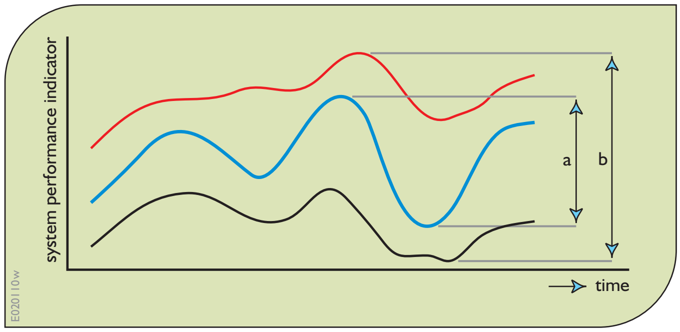



**Figure 9.4.** Time series of model output or system performance showing variability over time. Range 'a' results from the natural variability of input data over time. The extended range 'b' results from the variability of natural input data as well as from imprecision in input data measurement, parameter value estimation, model structure and errors in model solution algorithms. The extent of this range will depend on the confidence level associated with that range.

## 3.1. Natural Variability

The main source of hydrological model output value variability is the natural variability in hydrological and meteorological input series. Periods of normal precipitation and temperature can be interrupted by periods of extended drought and intense meteorological events such as hurricanes and tornadoes. There is no reason to believe that such events will not continue to occur and become even more frequent and extreme. Research has demonstrated that climate has been variable in the past and there are concerns about anthropogenic activities that may increase that variability each year. Sensitivity analysis can help assess the effect of errors in predictions if those predictions are based only on past records of historical time-series data describing precipitation, temperature and other exogenous forces across and on the border of the regions being studied.

Time-series input data are often actual, or at least based on, historical data. The time-series values typically describe historical conditions including droughts and wet periods. What is distinctive about natural uncertainty, as opposed to errors and uncertainty due to modelling limitations, is that natural variability in meteorological forces cannot be reduced by improving the model's structure, increasing the resolution of the simulation or better calibration of model parameters.

Errors result if meteorological values are not measured or recorded accurately, or if mistakes are made in the generation of computer data files. Furthermore, there is no assurance that the statistical properties of historical data will accurately represent those of future data. Actual future precipitation and temperature scenarios will be different from those in the past, and this difference in many cases may have a larger effect than the uncertainty due to incorrect parameter values. However, the effects of uncertainties in the parameter values used in stochastic generation models are often much more significant than the effects of using different stochastic generation models (Stedinger and Taylor, 1982).

While variability of model output is a direct result of variability of model input (e.g. hydrological and meteorological data), the extent of the variability, and its lower and upper limits, may also be affected by errors in the inputs, the values of parameters, initial boundary conditions, model structure, processes and solution algorithms.

Figure 9.4 illustrates the distinction between the variability of a system performance indicator due to input data variability, and the extended range of variability due to the total uncertainty associated with any combination of the causes listed in the previous section. This extended range is what is of interest to water resources planners and managers.

What can occur in practice is a time series of system performance indicator values that can range anywhere within or even outside the extended range, assuming the confidence level of that extended range is less than 100%. The confidence one can have that some future value of a time series will be within a given range is dependent on two factors. The first is the number of measurements used to compute the confidence limits. The second is the assumption that those measurements are representative (come from the same statistical or stochastic process yielding) future measurements, i.e. they come from the same statistical or stochastic process. Figure 9.5 illustrates this point.

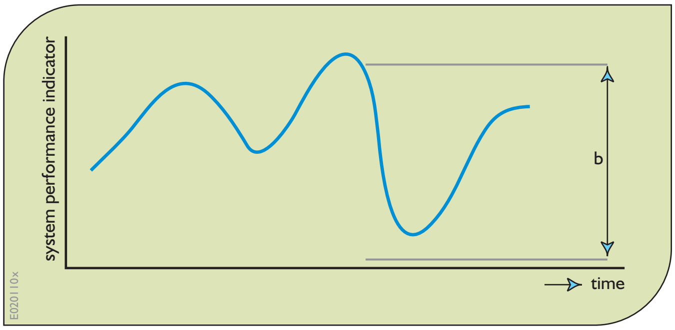



**Figure 9.5.** Typical time series of model output or system performance indicator values that are the result of input data variability and possible imprecision in input data measurement, parameter value estimation, model structure and errors in model solution algorithms.

Note that the time series may even contain values outside the range 'b' defined in Figure 9.4 if the confidence level of that range is less than 100%. Confidence intervals associated with less than 100% certainty will not include every possible value that might occur.

## 3.2. Knowledge Uncertainty

Referring to Figure 9.3, knowledge uncertainty includes model structure and parameter value uncertainties. First we consider parameter value uncertainty including boundary condition uncertainty, and then model and solution algorithm uncertainty.

### 3.2.1. Parameter Value Uncertainty

A possible source of uncertainty in model output results from uncertain estimates of various model parameter values. If the model calibration procedure were repeated using different data sets, different parameter values would result. Those values would yield different simulated system behaviour and, thus, different predictions. We can call this parameter uncertainty in the predictions because it is caused by imprecise parameter values. If such parameter value imprecision were eliminated, then the prediction would always be the same and so the parameter value uncertainty in the predictions would be zero. But this does not mean that predictions would be perfectly accurate.

In addition to parameter value imprecision, uncertainty in model output can result from imprecise specification of boundary conditions. These boundary conditions may be either fixed or variable. However, because they are not being computed on the basis of the state of the system, their values can be uncertain. These uncertainties can affect the model output, especially in the vicinity of the boundary, in each time step of the simulation.

### 3.2.2. Model Structural and Computational Errors

Uncertainty in model output can also result from errors in the model structure compared to the real system, and approximations made by numerical methods employed in the simulation. No matter how good our parameter value estimates, our models are not perfect and there is a residual model error. Increasing model complexity in order to more closely represent the complexity of the real system may not only add to the cost of data collection, but may also introduce even more parameters, and thus even more potential sources of error in model output. It is not an easy task to judge the appropriate level of model complexity and to estimate the resulting levels of uncertainty associated with various assumptions regarding model structure and solution methods. Kuczera (1988) provides an example of a conceptual hydrological modelling exercise with daily time steps where model uncertainty dominates parameter value uncertainty.

## 3.3. Decision Uncertainty

Uncertainty in model predictions can result from unanticipated changes in what is being modelled. These can include changes in nature, human goals, interests, activities, demands and impacts. An example of this is the deviation from standard or published operating policies by operators of infrastructure such as canal gates, pumps and reservoirs in the field, as compared to what is specified in documents and incorporated into the water systems models. Comparing field data with model data for model calibration may yield incorrect calibrations if operating policies actually implemented in the field differ significantly from those built into the models. What do operators do in times of stress?

What humans will want to achieve in the future may not be the same as what they want today. Predictions of people's future desires are clearly sources of uncertainty.

A perfect example of this can be seen in the very flat Greater Everglades region of south Florida in the United States. Fifty years ago, folk wanted the swampy region protected from floods and drained for agricultural and urban development. Today many want just the opposite, at least where there are no human settlements. They want to return to a more natural hydrological system with more wetlands and unobstructed flows, but now for ecological restoration reasons, which were not a major concern or much appreciated some half a century ago. Once the mosquitoes return, and if the sea level continues to rise, future populations who live there may want more flood control and drainage again. Who knows? Complex and changing social and economic processes influence human activities and their demands for water resources and environmental amenities over time. Some of these processes reflect changes in local concerns, interests and activities, but population migration and many economic activities and social attitudes can also reflect changing national and international trends.

Sensitivity scenarios that include human activities can help define the effects of those activities within an area. Careful attention should be given to the development of these alternative scenarios so that they capture the real forces or stresses that the system may face. The history of systems studies are full of examples where the issues studied were overwhelmed by much larger social forces resulting from, for example, the relocation of major economic activities, an oil embargo, changes in national demand for natural resources, economic recession, an act of terrorism or even war. One thing is certain: the future will be different than the past, and no one knows just how.

#### Surprises

Water resources managers may also want to consider how vulnerable a system is to undesirable environmental surprises. What havoc might an introduced species like the zebra mussel invading the Great Lakes of North America have in a particular watershed? Might some introduced disease suddenly threaten key plant or animal species? Might management plans have to be restructured to address the survival of species such as salmon in the Rhine River in Europe or in the Columbia River in North America? Such uncertainties are hard to anticipate when by their nature they will truly be surprises. But surprises should be expected. Hence system flexibility and adaptability should be sought to deal with changing management demands, objectives and constraints.

## 4. Sensitivity and Uncertainty Analyses

An uncertainty analysis is not the same as a sensitivity analysis. An uncertainty analysis attempts to describe the entire set of possible outcomes, together with their associated probabilities of occurrence. A sensitivity analysis attempts to determine the change in model output values that results from modest changes in model input values. A sensitivity analysis thus measures the change in the model output in a localized region of the space of inputs. However, one can often use the same set of model runs for both uncertainty analyses and sensitivity analyses. It is possible to carry out a sensitivity analysis of the model around a current solution, and then use it as part of a first-order uncertainty analysis.

This discussion begins by focusing on some methods of uncertainty analysis, then reviews various ways of performing and displaying sensitivity analyses.

### 4.1. Uncertainty Analyses

Recall that uncertainty involves the notion of randomness. If a value of a performance indicator or performance measure, like the phosphorus concentration or the depth of water at a particular location, varies, and this variation over space and time cannot be predicted with certainty, it is called a random variable. One cannot say with certainty what the value of a random variable will be but only the likelihood or probability that it will be within some specified range of values. The probabilities of observing particular ranges of values of a random variable are described or defined by a probability distribution. There are many types of distributions and each can be expressed in several ways as presented in Chapter 7.

Suppose the random variable is denoted as $X$. As discussed in Chapter 7, if the observed values of this random variable can only have discrete values, then the probability

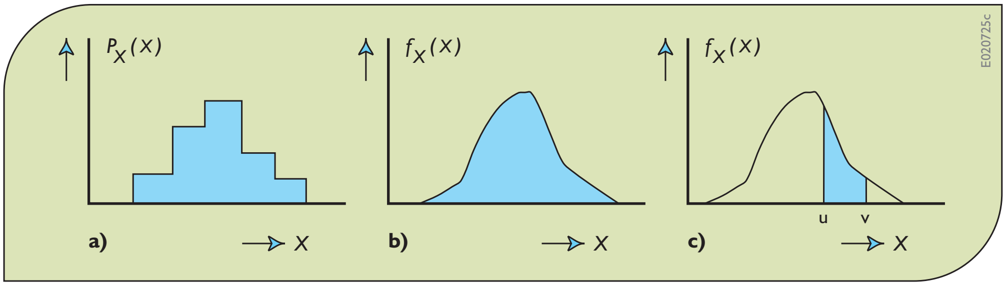



**Figure 9.6.** Probability distributions for a discrete or continuous random variable *X*. The area under the distributions (shaded areas in a and b) is 1, and the shaded area in c is the probability that the observed value *x* of the random variable *X* will be between *u* and *v*.

distribution of $X$ is easily described by a histogram, as shown in Figure 9.6a. The sum of the probabilities for all possible outcomes must equal 1. If the random variable is a continuous variable that can assume any real value over a range of values, the probability distribution of $X$ can be expressed as a continuous distribution, as shown in Figure 9.6b. The shaded area under the density function for the continuous distribution, is 1. The area between two values of the continuous random variable, such as between $u$ and $v$ in Figure 9.6c, represents the probability that the observed value $x$ of the random variable value $X$ will be within that range of values.

The probability distributions shown in Figure 9.6b and Figure 9.6c are called probability density functions (pdf) and are denoted by $f_X(x)$. The subscript $X$ on $P_X$ and $f_X$ represents the random variable, and the variable $x$ is some value of that random variable $X$.

Uncertainty analyses involve identifying characteristics of various probability distributions of model input and output variables, and subsequently functions of those random output variables that are performance indicators or measures. Often targets associated with these indicators or measures are themselves uncertain.

A complete uncertainty analysis would involve a comprehensive identification of all sources of uncertainty that contribute to the joint probability distributions of each input or output variable. Assume such analyses were performed for two alternative project plans, $A$ and $B$, and that the resulting probability density distributions for a specified performance measure were as shown in Figure 9.7. The figure also identifies the costs of these two projects. The introduction of two performance criteria – cost and probability of exceeding a performance measure target, e.g. a pollutant concentration standard – introduces a conflict where a tradeoff must be made.

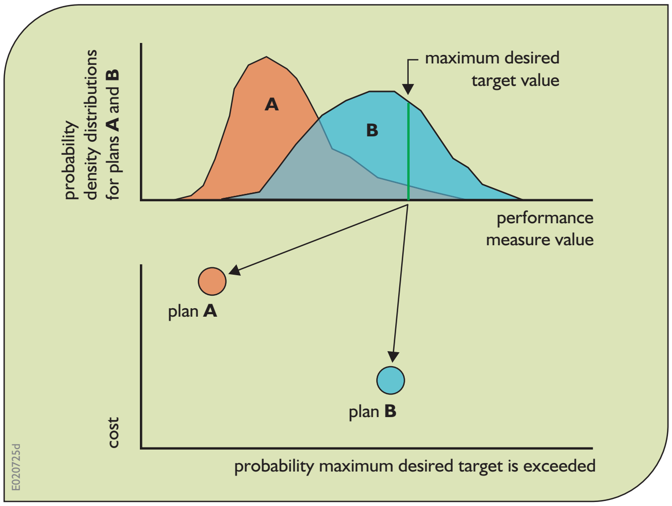



**Figure 9.7.** Tradeoffs involving cost and the probability that a maximum desired target value will be exceeded. In this illustration we want the lowest cost (*B* is best) and the lowest probability of exceedance (*A* is best).

### 4.1.1. Model and Model Parameter Uncertainties

Consider a situation as shown in Figure 9.8, in which, for a specific set of model inputs, the model outputs differ from the observed values, and for those model inputs, the observed values are always the same. Here, nothing occurs randomly. The model parameter values or model structure need to be changed. This is typically done in a model calibration process.

Given specific inputs, the outputs of deterministic models are always going to be the same each time those inputs are simulated. If for specified inputs to any simulation model the predicted output does not agree with the observed value, as shown in Figure 9.8, this could result from imprecision in the measurement of observed data. It could also result from imprecision in the model parameter values, the model structure or the algorithm used to solve the model.

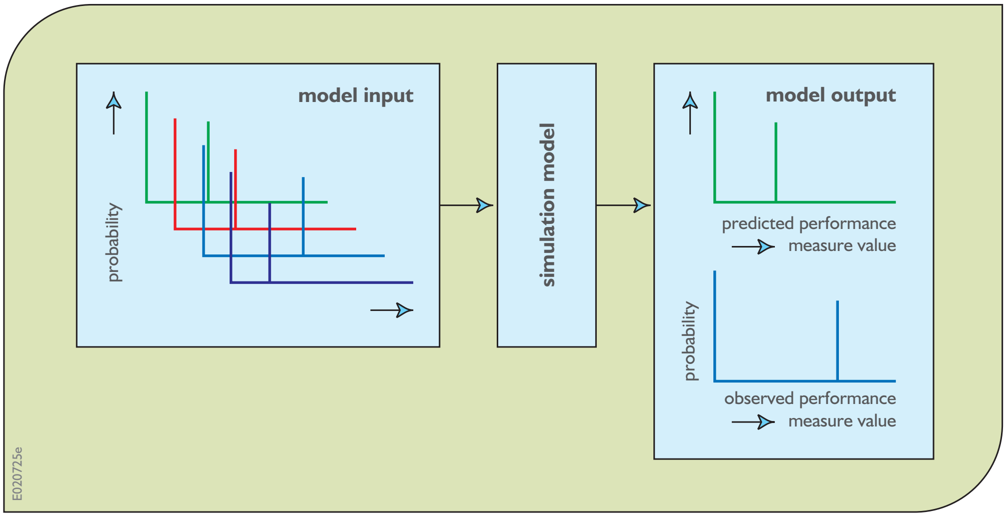



**Figure 9.8.** A deterministic system and a simulation model of that system needing calibration or modification in its structure. There is no randomness, only parameter value or model structure errors to be identified and corrected.

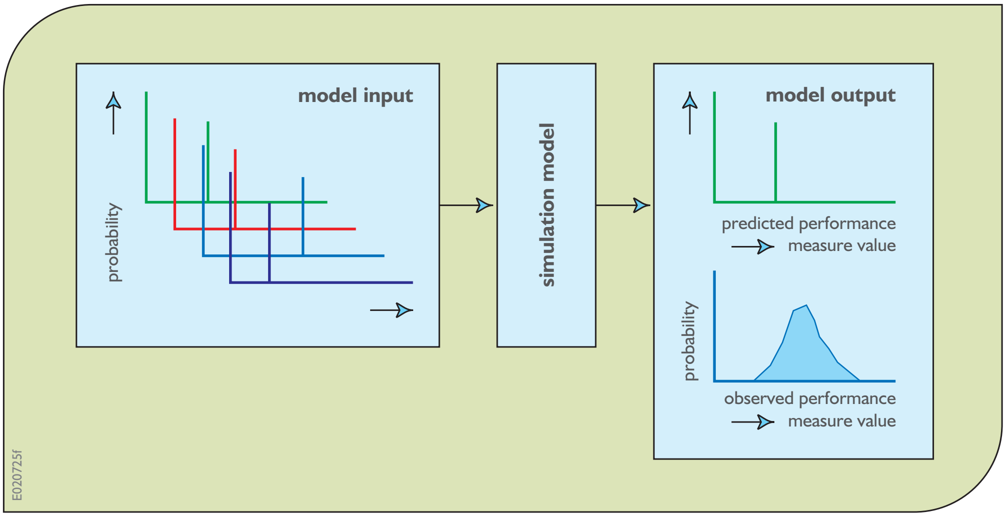



**Figure 9.9.** A deterministic simulation model of a 'random or stochastic' system. To produce the variability in the model output that is observed in the real system, even given the same input values, the model's parameter values may need to vary over distributions of values and/or the model structure may need modification along with additional model inputs.

Next, consider the same deterministic simulation model, but now assume at least some of the inputs are random (that is, not predictable), as may be case when random outputs of one model are used as inputs into another model. Random inputs will yield random outputs. The model input and output values can be described by probability distributions. If the uncertainty in the output is due only to the uncertainty in the input, the situation is similar to that shown in Figure 9.8. If the distribution of performance measure output values does not fit or is not identical to the distribution of observed performance measure values, then calibration of model parameter values or modification of the model structure may be needed.

If a model calibration or 'identification' exercise finds the 'best' values of the parameters to be outside reasonable ranges of values based on scientific knowledge, then the model structure or algorithm might be in error. Assuming that the algorithms used to solve the models are correct and that observed measurements of system performance vary for the same model inputs, as shown in Figure 9.9, it can be assumed that the model structure does not capture all the processes that are taking place that affect the value of the performance measures. This is often the case when relatively simple and low-resolution models are used to estimate the hydrological and ecological impacts of water and land management policies. However, even large and complex models can fail to include or adequately describe important phenomena.

In the presence of informational uncertainties, there may be considerable uncertainty about the values of the 'best' parameters during calibration. This problem becomes even more pronounced with increases in model complexity.

*An example:*

Consider the prediction of a pollutant concentration at some site downstream of a pollutant discharge site. Given a streamflow $Q$ (in units of $1000\,m^3$/day), the distance between the discharge site and the monitoring site, $X$(m), the pollutant decay rate constant $k$ ($day^{-1}$) and the pollutant discharge $W$ (Kg/day), we can use the following simple model to predict the concentration of the pollutant $C$ ($g/m^3 = mg/l$) at the downstream monitoring site:

$$C = (W/Q)\exp\{-k(X/U)\}$$

In the above equation, assume the velocity $U$ (m/day) is a known function of the streamflow $Q$.

In this case, the observed value of the pollutant concentration $C$ may differ from the computed value of $C$ even for the same inputs of $W$, $Q$, $k$, $X$ and $U$. Furthermore, this difference varies in different time periods. This apparent variability, as illustrated in Figure 9.9, can be simulated using the same model but by assuming a distribution of values for the decay rate constant $k$. Alternatively, the model structure can be modified to include the impact of streamflow temperature $T$ on the prediction of $C$.

$$C = (W/Q)\exp\{-k\theta^{T-20}(X/U)\}$$

Now there are two model parameters, the decay rate constant $k$ and the dimensionless temperature correction factor $\theta$, and an additional model input, the streamflow temperature, $T$, in °C. It could be that the variation in streamflow temperature was the major cause of the first equation's 'uncertainty' and that the assumed parameter distribution of $k$ was simply the result of the distribution of streamflow temperatures on the term $k\theta^{T-20}$.

If the output were still random given constant values of all the inputs, then another source of uncertainty exists. This uncertainty might be due to additional random loadings of the pollutant, possibly from non-point sources. Once again, the model could be modified to include these additional loadings if they are knowable. Assuming these additional loadings are not known, a new random parameter could be added to the input variable $W$ or to the right hand side of the equations above that would attempt to capture the impact on $C$ of these additional loadings. A potential problem, however, might be the likely correlation between those additional loadings and the streamflow $Q$.

While adding model detail removed some 'uncertainty' in the above example, increasing model complexity will not always eliminate or reduce uncertainty in model output. Adding complexity is generally not a good idea when the increased complexity is based on processes whose parameters are difficult to measure, when the right equations are not known at the scale of application, or when the amount of data for calibration is small compared to the number of parameters.

Even if more detailed models requiring more input data and more parameter values were to be developed, the likelihood of capturing all the processes occurring in a complex system is small. Hence, those involved will have to make decisions while taking this uncertainty into account. Imprecision will always exist due to a less than complete understanding of the system and the hydrological processes being modelled. A number of studies have addressed model simplification, but only in some simple cases have statisticians been able to identify just how one might minimize modelling related errors in model output values.

The problem of determining the 'optimal' level of modelling detail is particularly important when simulating the hydrological events at many sites over large areas. Perhaps the best approach for these simulations is to establish confidence levels for alternative sets of models and then statistically compare simulation results. But even this is not a trivial or cost-free task. Increases in the temporal or spatial resolution typically require considerable data collection and/or processing, model recalibrations, and possibly the solution of stability problems resulting from the numerical methods used in the models. Obtaining and implementing alternative hydrological simulation models will typically involve considerable investment of money and time for data preparation and model calibration.

What is needed is a way to predict the variability evident in the system shown in Figure 9.9. Instead of a fixed output vector for each fixed input vector, a distribution of outputs is needed for each performance measure based on fixed inputs (Figure 9.9) or a distribution of inputs (Figure 9.10.). Furthermore, the model output distribution for each performance measure should 'match' as well

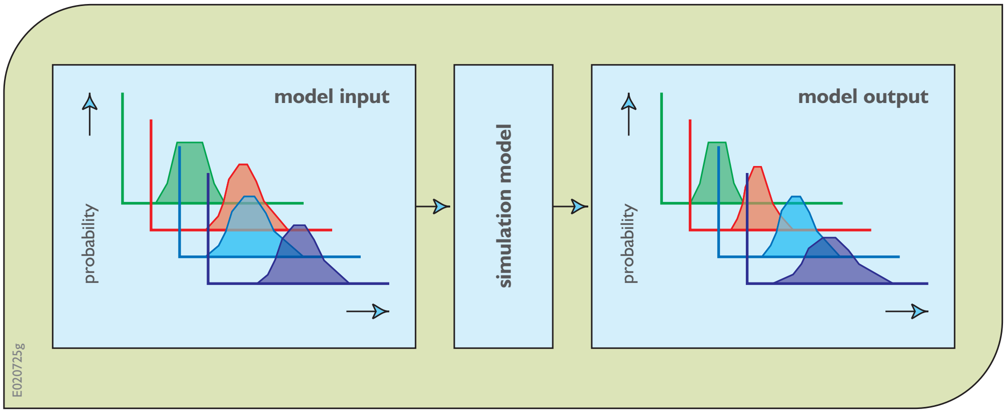



**Figure 9.10.** Simulating variable inputs to obtain probability distributions of predicted performance indices that match the probability distributions of observed performance values.

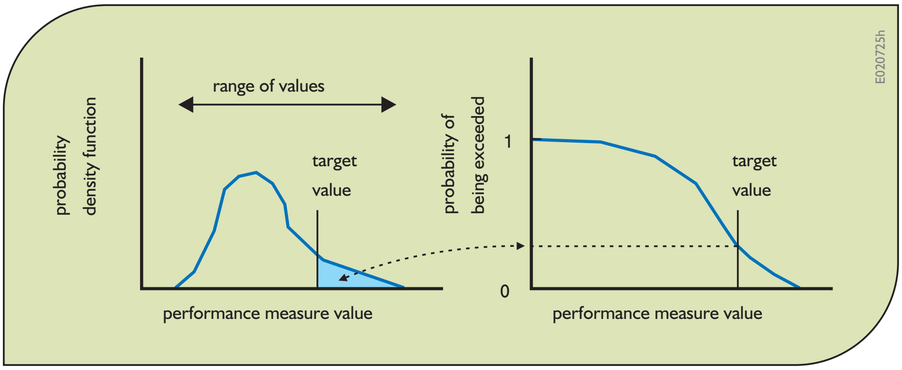



**Figure 9.11.** The distribution of performance measures defines a range of potential values and the likelihood that a specified target value will be exceeded. The shaded area under the density function on the left represents the probability that the target value will be exceeded. This probability is shown in the probability of exceedance plot on the right.

as possible the observed distribution of that performance measure.

### 4.1.2. What Uncertainty Analysis Can Provide

An uncertainty analysis takes a set of randomly chosen input values (which can include parameter values), passes them through a model (or transfer function) to obtain the distributions (or statistical measures of the distributions) of the resulting outputs. As illustrated in Figure 9.11, the output distributions can be used to

- describe the range of potential outputs of the system at some probability level
- estimate the probability that the output will exceed a specific threshold or performance measure target value.

Uncertainty analyses are often used to make general inferences, such as the following:

- estimating the mean and standard deviation of the outputs
- estimating the probability the performance measure will exceed a specific threshold
- assigning a reliability level to a function of the outputs, for example, the range of function values that is likely to occur with some probability
- describing the likelihood of different potential outputs of the system
- estimating the relative impacts of input variable uncertainties.

Implicit in any uncertainty analysis are the assumptions that statistical distributions for the input values are correct and that the model is a sufficiently realistic description of the processes taking place in the system. Neither of these assumptions is likely to be entirely correct.

## 4.2. Sensitivity Analyses

'Sensitivity analysis' aims to describe how much model output values are affected by changes in model input values. It is the investigation of the importance of imprecision or uncertainty in model inputs in a decision-making or modelling process. The exact character of a sensitivity analysis depends upon the particular context and the questions of concern. Sensitivity studies can provide a general assessment of model precision when used to assess system performance for alternative scenarios, as well as detailed information addressing the relative significance of errors in various parameters. As a result, sensitivity results should be of interest to the general public, federal and state management agencies, local water resources planners and managers and model users and developers.

Clearly, upper level management and the public may be interested in more general statements of model precision, and should be provided such information along with model predictions. On the other hand, detailed studies addressing the significance and interactions among individual parameters would likely be meaningful to model developers and some model users. They can use such data to interpret model results and to identify to where their efforts to improve models and input values should be directed.

Initial sensitivity analysis studies could focus on two products:

- detailed results that guide research and assist model development efforts
- calculation of general descriptions of uncertainty associated with model predictions, so that policy decisions can reflect both the modelling efforts' best prediction of system performance and the precision of such predictions.

In the first case, knowing the relative uncertainty in model projections due to possible errors in different sets of parameters and input data should assist in efforts to improve the precision of model projections. This knowledge should also contribute to a better understanding of the relationships between model assumptions, parameters, data and model predictions.

In the second case, knowing the relative precision associated with model predictions should have a significant effect on policy development. For example, the analysis may show that, given data inadequacies, there are very large error bands associated with some model variables. When such large uncertainties exist, predictions should be used with appropriate scepticism. Incremental strategies should be explored along with monitoring so that greater experience can accumulate to resolve some of these uncertainties.

Sensitivity analysis features are available in many linear and non-linear programming (optimization) packages. They identify the changes in the values of the objective function and unknown decision-variables given a change in the model input values, and a change in levels set for various constraints (Chapter 4). Thus, sensitivity analysis can address the change in 'optimal' system performance associated with changes in various parameter values, and also how 'optimal' decisions would change with changes in resource constraint levels or target output requirements. This kind of sensitivity analysis provides estimates of how much another unit of resource would be worth, or what 'cost' a proposed change in a constraint places on the optimal solution. This information is of value to those making design decisions.

Various techniques have been developed to determine how sensitive model outputs are to changes in model inputs. Most approaches examine the effects of changes in a single parameter value or input variable assuming no changes in all the other inputs. Sensitivity analyses can be extended to examine the combined effects of multiple sources of error, as well.

Changes in particular model input values can affect model output values in different ways. It is generally true that only a relatively few input variables dominate or substantially influence the values of a particular output variable or performance indicator at a particular location and time. If the range of uncertainty of only some of the output data is of interest, then undoubtedly only those input data that significantly affect the values of those output data need be included in the sensitivity analysis.

If input data estimates are based on repeated measurements, a frequency distribution can be estimated that characterizes natural variability. The shorter the record of measurements, the greater will be the uncertainty regarding the long-term statistical characteristics of that variability. If obtaining a sufficient number of replicate measurements is not possible, subjective estimates of input data ranges and probability distributions are often made. Using a mixture of subjective estimates and actual measurements does not affect the application of various sensitivity analysis methods that can use these sets or distributions of input values, but it may affect the conclusions that can be drawn from the results of these analyses.

It would be nice to have available accurate and easy-to-use analytical methods for relating errors in input data to errors in model outputs, and to errors in system performance indicator values that are derived from model output. Such analytical methods do not exist for complex simulation models. However, methods based on simplifying assumptions and approximations can be used to yield useful sensitivity information. Some of these are reviewed in the remainder of this chapter.

### 4.2.1. Sensitivity Coefficients

One measure of sensitivity is the sensitivity coefficient. This is the derivative of a model output variable with respect to an input variable or parameter. A number of sensitivity analysis methods use these coefficients. First-order and approximate first-order sensitivity analyses are two such methods that will be discussed later. Analytical methods are faced with considerable difficulties in:

- obtaining the derivatives for many models
- needing to assume mathematical (usually linear) relationships when obtaining estimates of derivatives by making small changes of input data values near their nominal or most likely values
- having large variances associated with most hydrological process models.

These have motivated the replacement of analytical methods by numerical and statistical approaches to sensitivity analysis.

Implicit in any sensitivity analysis are the assumptions that statistical distributions for the input values are correct and that the model is a sufficiently realistic description of the processes taking place in the system. Neither of these assumptions is likely to be entirely correct.

The importance of the assumption that the statistical distributions for the input values are correct is easy to check by using different distributions for the input parameters. If the outputs vary significantly, then the output is sensitive to the specification of the input distributions, and hence they should be defined with care. A relatively simple deterministic sensitivity analysis can be of value here (Benaman, 2002). A sensitivity coefficient can be used to measure the magnitude of change in an output variable $Q$ per unit change in the magnitude of an input parameter value $P$ from its base value $P_0$. Let $SI_{PQ}$ be the sensitivity index for an output variable $Q$ with respect to a change $\Delta P$ in the value of the input variable $P$ from its base value $P_0$. Noting that the value of the output $Q(P)$ is a function of $P$, the sensitivity index could be defined as

$$SI_{PQ} = [Q(P_0 + \Delta P) - Q(P_0 - \Delta P)]/2\Delta P \tag{9.1}$$

Other sensitivity indices could be defined (McCuen, 1973). Letting the index $i$ represent a decrease and $j$ represent an increase in the parameter value from its base value $P_0$, the sensitivity index $SI_{PQ}$ for parameter $P$ and output variable $Q$ could be defined as

$$SI_{PQ} = \{|(Q_0 - Q_i)/(P_0 - P_i)| + |(Q_0 - Q_j)/(P_0 - P_j)|\}/2 \tag{9.2}$$

or

$$SI_{PQ} = \max\{|(Q_0 - Q_i)/(P_0 - P_i)|, |(Q_0 - Q_j)/(P_0 - P_j)|\} \tag{9.3}$$

A dimensionless expression of sensitivity is the elasticity index, $EI_{PQ}$, which measures the relative change in output $Q$ for a relative change in input $P$, and could be defined as

$$EI_{PQ} = [P_0/Q(P_0)]SI_{PQ} \tag{9.4}$$

### 4.2.2. A Simple Deterministic Sensitivity Analysis Procedure

This deterministic sensitivity analysis approach is very similar to those most often employed in the engineering economics literature. It is based on the idea of varying one uncertain parameter value, or set of parameter values, at a time. The ideas are applied to a water quality example to illustrate their use.

The output variable of interest can be any performance measure or indicator. Thus, one does not know if more or less of a given variable is better or worse. Perhaps too much and/or too little is undesirable. The key idea is that, whether employing physical measures or economic metrics of performance, various parameters (or sets of associated parameters) are assigned high and low values. Such ranges may reflect either the differences between the minimum and maximum values for each parameter, the 5$^{th}$ and 95$^{th}$ percentiles of a parameter's distribution, or points corresponding to some other criteria. The system model is then run with the various alternatives, one at a time, to evaluate the impact of those errors in various sets of parameter values on the output variable.

Table 9.1 illustrates the character of the results that one would obtain. Here $Y_0$ is the nominal value of the model output when all parameters assume the estimated best values, and $Y_{i,L}$ and $Y_{i,H}$ are the values obtained by increasing or decreasing the values of the $i^{th}$ set of parameters.

A simple water quality example is employed to illustrate this deterministic approach to sensitivity analysis. The analysis techniques illustrated here are just as applicable to complex models. The primary difference is that more work would be required to evaluate the various alternatives with a more complex model, and the model responses might be more complicated.

The simple water quality model is provided by Vollenweider's empirical relationship for the average phosphorus concentration in lakes (Vollenweider, 1976). He found that the phosphorus concentration, $P$ (mg/m$^3$), is a function of the annual phosphorus loading rate, $L$ (mg/m$^2 \cdot$ a), the annual hydraulic loading, $q$ (m/a, or more exactly, m$^3$/m$^2 \cdot$ a), and the mean water depth, $z$ (m):

$$P = (L/q)/[1 + (z/q)^{0.5}] \tag{9.5}$$

$L/q$ and $P$ have the same units; the denominator is an empirical factor that compensates for nutrient recycling and elimination within the aquatic lake environment.

| parameter set | low value | nominal | high value |
|---|---|---|---|
| 1 | $Y_{2,L}$ | $Y_0$ | $Y_{1,H}$ |
| 2 | $Y_{2,L}$ | $Y_0$ | $Y_{2,H}$ |
| 3 | $Y_{3,L}$ | $Y_0$ | $Y_{3,H}$ |
| 4 | $Y_{4,L}$ | $Y_0$ | $Y_{4,H}$ |

E020829m

**Table 9.1.** Sensitivity of model output $Y$ to possible errors in four parameter sets containing a single parameter or a group of parameters that vary together.

Data for Lake Ontario in North America would suggest that reasonable values of the parameters are $L = 680$ mg/m$^3$, $q = 10.6$ m/a, and $z = 84$ m, yielding $P = 16.8$ mg/m$^3$. Values of phosphorus concentrations less than 10 mg/m$^3$ are considered oligotrophic, whereas values greater than 20 mg/m$^3$ generally correspond to eutrophic conditions. Reasonable ranges reflecting possible errors in the three parameters yield the values in Table 9.2.

One may want to display these results so they can be readily visualized and understood. A tornado diagram (Eschenbach, 1992) would show the lower and upper values of $P$ obtained from variation of each parameter, with the parameter with the widest limits displayed on top, and the parameter with smallest limits on the bottom. Tornado diagrams (Figure 9.12) are easy to construct and can include a large number of parameters without becoming crowded.

An alternative to tornado diagrams is a Pareto chart showing the width of the uncertainty range associated with each variable, ordered from largest to smallest. A Pareto chart is illustrated in Figure 9.13.

Another visual presentation is a spider plot showing the impact of uncertainty in each parameter on the variable in question, all on the same graph (Eschenback, 1992; DeGarmo et al., 1993). A spider plot, Figure 9.14, shows the particular functional response of the output to each parameter on a common scale, so one needs a common metric to represent changes in all of the parameters. Here, we use percentage change from the nominal or best values.

Spider plots are a little harder to construct than tornado diagrams, and can generally include only four or five variables without becoming crowded. However, they provide a more complete view of the relationships between each parameter and the performance measure. In particular, a spider plot reveals non-linear relationships

**Table 9.2.** Sensitivity of estimates of phosphorus concentration (mg/m$^3$) to model parameter values. The two right-most values in each row correspond to the low and high values of the parameter, respectively.

| | parameter value | | phosphorus concentration | |
|---|---|---|---|---|
| | low | high | $P_{low}$ | $P_{high}$ |
| **$L$ — $P$ loading (mg/m$^3$. a)** | 500 | 900 | 12.4 | 22.3 |
| **$q$ — hydraulic loading (m/a)** | 8 | 13.5 | 20.0 | 14.4 |
| **$z$ — mean depth (m)** | 81 | 87 | 17.0 | 16.6 |

E020829n

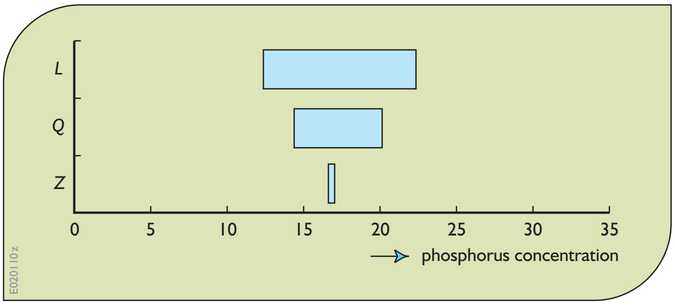



**Figure 9.12.** A tornado diagram showing the range of the output variables representing phosphorus concentrations for high and low values of each of the parameter sets. Parameters are sorted so that the largest range is on top, and the smallest on the bottom, so the diagram looks like a tornado.

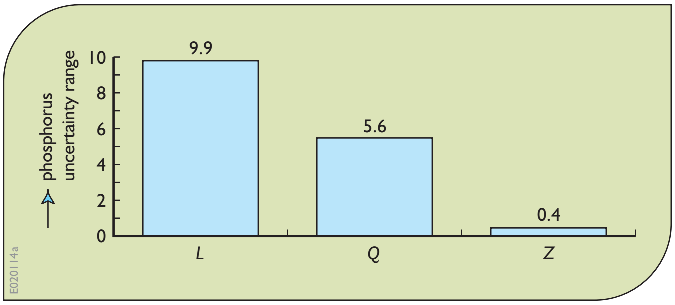



**Figure 9.13.** A Pareto chart showing the range of the output variable representing phosphorus concentrations resulting from high and low values of each parameter set considered.

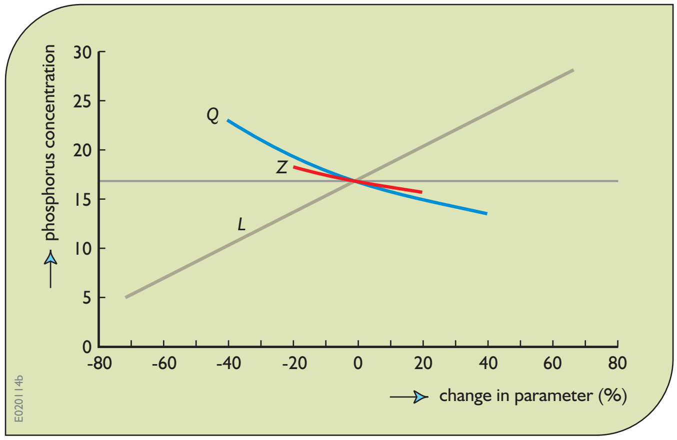



**Figure 9.14.** A spider Plot illustrates the relationships between model output describing phosphorus concentrations and variations in each of the parameter sets, expressed as a percentage deviation from their nominal values.

and the relative sensitivity of the performance measure to (percentage) changes in each variable.

In the spider plot, the linear relationship between $P$ and $L$ and the gentle non-linear relationship between $P$ and $q$ is illustrated. The range for $z$ has been kept small, given the limited uncertainty associated with that parameter.

### 4.2.3. Multiple Errors and Interactions

An important issue that should not be ignored is the impact of simultaneous errors in more than one parameter. Probabilistic methods directly address the occurrence of simultaneous errors, but the correct joint distribution needs to be employed. With simple sensitivity analysis procedures, errors in parameters are generally investigated one at a time or in groups. The idea of considering pairs or sets of parameters is discussed here.

*Groups of factors*

It is often the case that reasonable error scenarios would have several parameters changing together. For this reason, the alternatives have been called parameter sets. For example, possible errors in water depth would be accompanied by corresponding variations in aquatic vegetation and chemical parameters. Likewise, alternatives related to changes in model structure might be accompanied with variations in several parameters. In other cases, there may be no causal relationship among possible errors (such as model structure versus inflows at the boundary of the modelled region), but they might still interact to affect the precision of model predictions.

*Combinations*

If one or more non-grouped parameters interact in significant ways, then combinations of one or more errors should be investigated. However, one immediately runs into a combinatorial problem. If each of $m$ parameters can have 3 values (high, nominal, and low), then there are $3^m$ combinations, as opposed to $2m + 1$ if each parameter is varied separately. (For $m = 5$, the differences are $3^5 = 243$ versus $2(5) + 1 = 11$.) These numbers can be reduced by considering instead only combinations of extremes so that only $2^m + 1$ cases need be considered ($2^5 + 1 = 33$), which is a more manageable number. However, all of the parameters would be at one extreme or the other, and such situations would be very unusual.


**Figure 9.15.** Pareto diagram showing errors in phosphorus concentrations for all combinations of pairs of input parameters errors. A '+' indicates a high value, and a '−' indicates a low value for indicated parameter. *L* is the phosphorus loading rate, *q* is the hydraulic loading, and *z* is the mean lake depth.

*Two factors at a time*

A compromise is to consider all pairs of two parameters at a time. There are $m(m-1)/2$ possible pairs of $m$ parameters. Each parameter has a high and low value. Since there are 4 combinations of high and low values for each pair, there are a total of $2m(m-1)$ combinations. (For $m = 5$ there are 40 combinations of two parameters each having two values.)

The presentation of these results could be simplified by displaying for each case only the maximum error, which would result in $m(m-1)/2$ cases that might be displayed in a Pareto diagram. This would allow identification of those combinations of two parameters that might yield the largest errors and thus are of most concern.

For the water quality example, if one plots the absolute value of the error for all four combinations of high $(+)$ and low $(-)$ values for each pair of parameters, they obtain Figure 9.15.

Considering only the worst error for each pair of variables yields Figure 9.16.

Here we see, as is no surprise, that the worst error results from the most unfavourable combination of $L$ and $q$ values. If both parameters have their most unfavourable values, the predicted phosphorus concentration would be 27 mg/m$^3$.

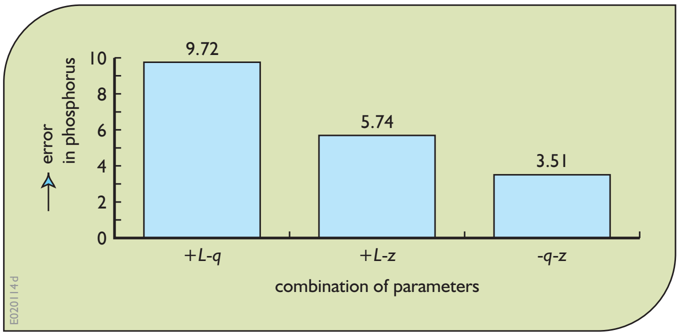



**Figure 9.16.** Pareto diagram showing worst error combinations for each pair of input parameters. A '+' indicates a high value, and a '−' indicates a low value for indicated parameter.

### 4.2.4. First-Order Sensitivity Analysis

The above deterministic analysis has trouble representing reasonable combinations of errors in several parameter sets. If the errors are independent, it is highly unlikely that any two sets would actually be at their extreme ranges at the same time. By defining probability distributions of the values of the various parameter sets, and specifying their joint distributions, a probabilistic error analysis can be conducted. In particular, for a given performance indicator, one can use multivariate linear analyses to evaluate the approximate impact on the performance indices of uncertainty in various parameters. As shown below, the impact depends upon the square of the sensitivity coefficients (partial derivatives) and the variances and covariances of the parameter sets.

For a performance indicator $I = F(Y)$, which is a function $F(\cdot)$ of model outputs $Y$, which are in turn a function $g(P)$ of input parameters $P$, one can use a multivariate Taylor series approximation of $F$ to obtain the expected value and variance of the indicator:

$$E[I] = F(\text{based on mean values of input parameters}) + (1/2)\left\{\sum_i \sum_j [\partial F^2/\partial P_i \partial P_j]\ \mathrm{Cov}[P_i, P_j]\right\} \qquad (9.6)$$

and

$$\mathrm{Var}[I] = \sum_i \sum_j (\partial F/\partial P_i)(\partial F/\partial P_j)\ \mathrm{Cov}[P_i, P_j] \tag{9.7}$$

where $(\partial F/\partial P_i)$ are the partial derivatives of the function $F$ with respect to $P_i$ evaluated at the mean value of the input parameters $P_i$, and $\partial F^2/\partial P_i \partial P_j$ are the second partial derivatives. The covariance of two random input parameters $P_i$ and $P_j$ is the expected value of the product of differences between the values and their means:

$$\mathrm{Cov}[P_i, P_j] = E[(P_i - E[P_i])(P_j - E[P_j])] \tag{9.8}$$

If all the parameters are independent of each other, and the second-order terms in the expression for the mean $E[I]$ are neglected, one obtains

$$E[I] = F(\text{based on mean values of input parameters}) \tag{9.9}$$

and

$$\mathrm{Var}[I] = \sum_i [\partial F/\partial P_i]^2\ \mathrm{Var}[P_i] \tag{9.10}$$

(Benjamin and Cornell, 1970). Equation 9.6 for $E[I]$ shows that in the presence of substantial uncertainty, the mean of the output from non-linear systems is not simply the system output corresponding to the mean of the parameters (Gaven and Burges, 1981). This is true for any non-linear function.

Of interest in the analysis of uncertainty is the approximation for the variance $\mathrm{Var}[I]$ of indicator $I$. In Equation 9.10 the contribution of $P_i$ uncertainty to the variance of $I$ equals $\mathrm{Var}[P_i]$ times $[\partial F/\partial P_i]^2$, which are the squares of the sensitivity coefficients for indicator $I$ with respect to each input parameter value $P_i$.

## An Example of First-Order Sensitivity Analysis

It may appear that first-order analysis is difficult because the partial derivatives of the performance indicator $I$ are needed with respect to the various parameters. However, reasonable approximations of these sensitivity coefficients can be obtained from the simple sensitivity analysis described in Table 9.3. In that table, three different parameter sets, $P_i$, are defined in which one parameter of the set is at its high value, $P_{iH}$, and one is at its low value, $P_{iL}$, to produce corresponding values (called high, $I_{iH}$, and low, $I_{iL}$) of a system performance indicator $I$.

E020829p

| parameter set | value: low | value: high | sensitivity coefficient |
|---|---|---|---|
| 1 | $I_{1L}$ | $I_{1H}$ | $[I_{1H}-I_{1L}]/[P_{1H}-P_{1L}]$ |
| 2 | $I_{2L}$ | $I_{2H}$ | $[I_{2H}-I_{2L}]/[P_{2H}-P_{2L}]$ |
| 3 | $I_{3L}$ | $I_{3H}$ | $[I_{3H}-I_{3L}]/[P_{3H}-P_{3L}]$ |

**Table 9.3.** Approximate parameter sensitivity coefficients.

It is then necessary to estimate some representation of the variances of the various parameters with some consistent procedure. For a normal distribution, the distance between the $5^{th}$ and $95^{th}$ percentiles is 1.645 standard deviations on each side of the mean, or 2(1.645) = 3.3 standard deviations. Thus, if the high/low range is thought of as approximately a 5–95 percentile range for a normally distributed variate, a reasonable approximation of the variance might be

$$\mathrm{Var}[P_i] = \{[P_{iH} - P_{iL}]/3.3\}^2 \tag{9.11}$$

This is all that is needed. Use of these average sensitivity coefficients is very reasonable for modelling the behaviour of the system performance indicator $I$ over the indicated ranges.

As an illustration of the method of first-order uncertainty analysis, consider the lake-quality problem described earlier. The 'system performance indicator' in this case is the model output, the phosphorus concentration $P$, and the input parameters, now denoted as $X = L$, $q$ and $z$. The standard deviation of each parameter is assumed to be the specified range divided by 3.3. Average sensitivity coefficients $\partial P/\partial X$ were calculated. The results are reported in Table 9.4.

Assuming the parameter errors are independent:

$$\mathrm{Var}[P] = 9.18 + 2.92 + 0.02 = 12.12 \tag{9.12}$$

The square root of 12.12 is the standard deviation and equals 3.48. This agrees well with a Monte Carlo analysis reported below.

Note that 100*(9.18/12.12), or about 76% of the total parameter error variance in the phosphorus concentration $P$, is associated in the phosphorus loading rate $L$ and the remaining 24% is associated with the hydrological loading $q$. Eliminating the uncertainty in $z$ would have a negligible impact on the overall model error. Likewise, reducing the

**Table 9.4.** Calculation of approximate parameter sensitivity coefficients.

| variable $X$ | units | $\partial P/\partial X$ | St Dev[X] | $(\partial P/\partial X)^2$ Var[X] | % |
|---|---|---|---|---|---|
| $L$ | mg/m$^2$.a | 0.025 | 121.21 | 9.18 | 75.7 |
| $q$ | m/a | -1.024 | 1.67 | 2.92 | 24.1 |
| $z$ | m | -0.074 | 1.82 | 0.02 | 0.2 |

E020829q

error in $q$ would at best have a modest impact on the total error.

Due to these uncertainties, the estimated phosphorus concentration has a standard deviation of 3.48. Assuming the errors are normally distributed, and recalling that $\pm$ 1.645 standard deviations around the mean define a 5–95 percentile interval, the 5–95 percentile interval would be about

$$16.8 \pm 1.645(3.48)\ \text{mg/m}^3 = 16.8 \pm 5.7\ \text{mg/m}^3$$
$$= 11.1 \text{ to } 22.5\ \text{mg/m}^3 \qquad (9.13)$$

These error bars indicate that there is substantial uncertainty associated with the phosphorus concentration $P$, primarily due to uncertainty in the loading rate $L$.

The upper bound of 22.6 mg/m$^3$ is considerably less than the 27 mg/m$^3$ that would be obtained if both $L$ and $q$ had their most unfavourable values. In a probabilistic analysis with independent errors, such a combination is highly unlikely.

### Warning on Accuracy

First-order uncertainty analysis is indeed an approximate method based upon a linearization of the response function represented by the full simulation model. It may provide inaccurate estimates of the variance of the response variable for non-linear systems with large uncertainty in the parameters. In such cases, Monte Carlo simulation (discussed below and in Chapters 7 and 8) or the use of higher-order approximation may be required. Beck (1987) cites studies that found that Monte Carlo and first-order variances were not appreciably different, and a few studies that found specific differences. Differences are likely to arise when the distributions used for the parameters are bimodal (or otherwise unusual), or some rejection algorithm is used in the Monte Carlo analysis to exclude some parameter combinations. Such errors can result in a distortion in the ranking of predominant sources of uncertainty. However, in most cases very similar results were obtained.

### 4.2.5. Fractional Factorial Design Method

An extension of first-order sensitivity analysis would be a more complete examination of the response surface using a careful statistical design. First, consider a complete factorial design. Input data are divided into discrete 'levels'. The simplest case is two levels, which can be defined as a nominal value, and a high (low) value. Simulation runs are made for all combinations of parameter levels. For $n$ different inputs, this would require $2^n$ simulation runs. Hence, for a three-input variable or parameter problem, eight runs would be required. If four discrete levels of each input variable or parameter were allowed to provide a more reasonable description of a continuous variable, the three-input data problem would require $4^3$ or 64 simulation runs. Clearly this is not a useful tool for large regional water resources simulation models.

A fractional factorial design involves simulating only a fraction of what is required from a full factorial design method. The loss of information prevents a complete analysis of the impacts of each input variable or parameter on the output.

To illustrate the fractional factorial design method, consider the two-level with three-input variable or parameter problem. Table 9.5 below shows the eight simulations required for a full factorial design method. The '+' and '−' symbols show the upper and lower levels of each input variable or parameter $P_i$ where $i = 1, 2, 3$. If all eight simulations were performed, seven possible effects could be estimated. These are the individual effects of the three inputs $P_1$, $P_2$, and $P_3$; the three two-input variable or

E020829r

| simulation run | value of input $P_1$ | $P_2$ | $P_3$ | value of output - variable Y |
|---|---|---|---|---|
| 1 | - | - | - | $Y_1$ |
| 2 | + | - | - | $Y_2$ |
| 3 | - | + | - | $Y_3$ |
| 4 | + | + | - | $Y_4$ |
| 5 | - | - | + | $Y_5$ |
| 6 | + | - | + | $Y_6$ |
| 7 | - | + | + | $Y_7$ |
| 8 | + | + | + | $Y_8$ |

**Table 9.5.** A three-input factorial design.

parameter interactions, $(P_1)(P_2)$, $(P_1)(P_3)$, and $(P_2)(P_3)$; and the one three-input variable or parameter interaction, $(P_1)(P_2)(P_3)$.

Consider an output variable $Y$, where $Y_j$ is the value of $Y$ in the $j^{\text{th}}$ simulation run. Then an estimate of the effect, denoted $\delta(Y \mid P_i)$, that input variable or parameter $P_i$ has on the output variable $Y$ is the average of the four separate effects of varying $P_i$:

For $i = 1$:

$$\delta(Y \mid P_1) = 0.25[(Y_2 - Y_1) + (Y_4 - Y_3) + (Y_6 - Y_5) + (Y_8 - Y_7)] \quad (9.14)$$

Each difference in parentheses is the difference between a run in which $P_1$ is at its upper level and a run in which $P_1$ is at its lower level, but the other two parameter values, $P_2$ and $P_3$, are unchanged. If the effect is equal to 0, then, in that case, $P_1$ has on average no impact on the output variable $Y$.

Similarly the effects of $P_2$ and $P_3$, on variable $Y$ can be estimated as:

$$\delta(Y \mid P_2) = 0.25\{(Y_3 - Y_1) + (Y_4 - Y_2) + (Y_7 - Y_5) + (Y_8 - Y_6)\} \quad (9.15)$$

and

$$\delta(Y \mid P_3) = 0.25\{(Y_5 - Y_1) + (Y_6 - Y_2) + (Y_7 - Y_3) + (Y_8 - Y_4)\} \quad (9.16)$$

Consider next the interaction effects between $P_1$ and $P_2$. This is estimated as the average of the difference between the average $P_1$ effect at the upper level of $P_2$ and the average $P_1$ effect at the lower level of $P_2$. This is the same as the difference between the average $P_2$ effect at the upper level of $P_1$ and the average $P_2$ effect at the lower level of $P_1$:

$$\begin{aligned}\delta(Y \mid P_1, P_2) &= (1/2)\{[(Y_8 - Y_7) + (Y_4 - Y_3)]/2 \\ &\quad - [(Y_2 - Y_1) + (Y_6 - Y_5)]/2\} \\ &= (1/4)\{[(Y_8 - Y_6) + (Y_4 - Y_2)] \\ &\quad - [(Y_3 - Y_1) + (Y_7 - Y_5)]\} \end{aligned} \quad (9.17)$$

Similar equations can be derived to show the interaction effects between $P_1$ and $P_3$, and between $P_2$ and $P_3$ and the interaction effects among all three inputs $P_1$, $P_2$, and $P_3$.

Now assume only half of the simulation runs were performed, perhaps runs 2, 3, 5 and 8 in this example. If only outputs $Y_2$, $Y_3$, $Y_5$, and $Y_8$ are available, for our example:

$$\begin{aligned}\delta(Y \mid P_3) &= \delta(Y \mid P_1, P_2) \\ &= 0.5\{(Y_8 - Y_3) - (Y_2 - Y_5)\}\end{aligned} \quad (9.18)$$

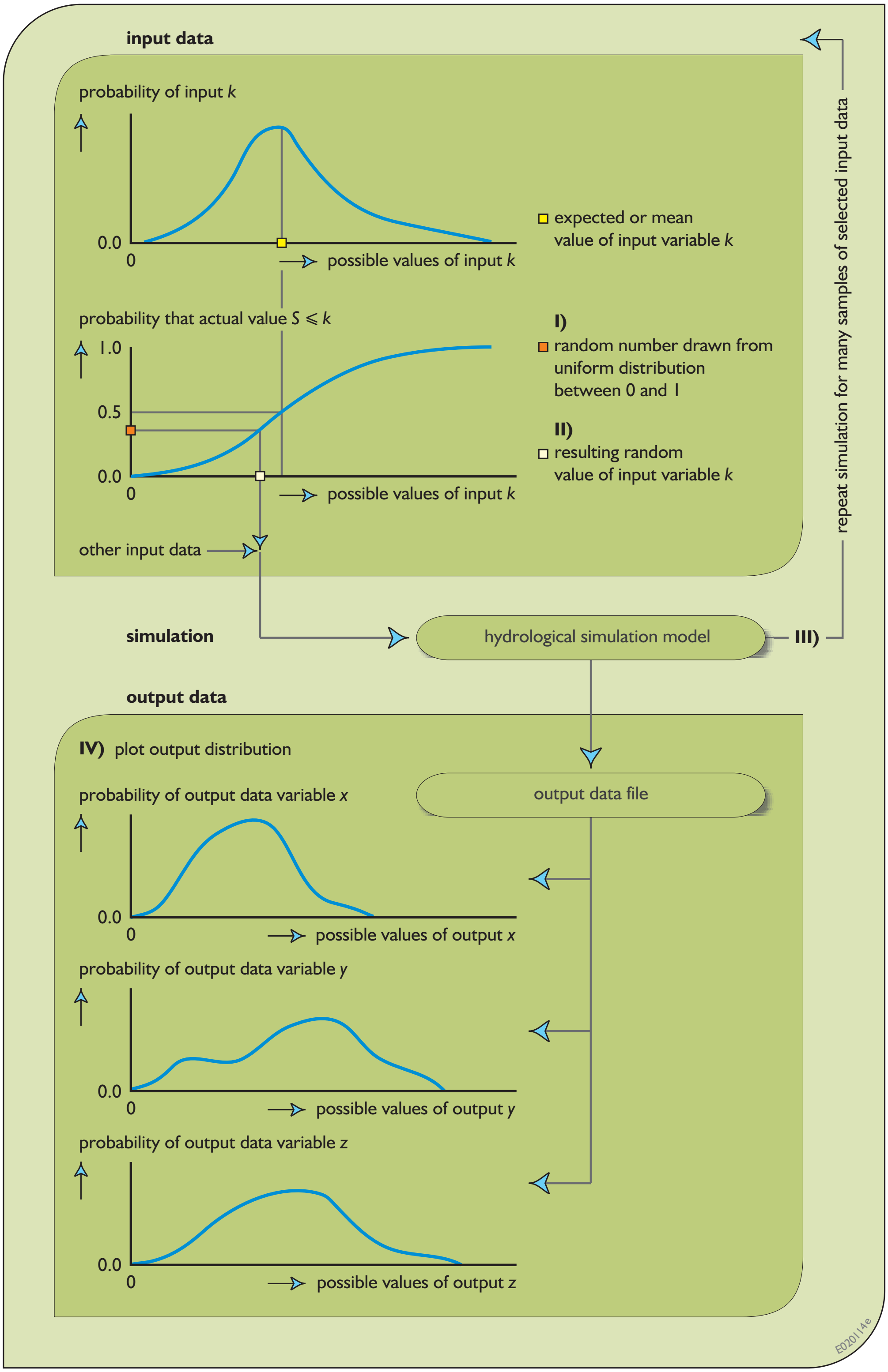
The separate effects of $P_3$ and of $P_1P_2$ are not available from the output. This is the loss in information resulting from fractional instead of complete factorial design.

### 4.2.6. Monte Carlo Sampling Methods

The Monte Carlo method of performing sensitivity analyses, illustrated in Figure 9.17, first selects a random set of input data values drawn from their individual probability distributions. These values are then used in the simulation model to obtain some model output variable values. This process is repeated many times, each time making sure the model calibration is valid for the input data values chosen. The end result is a probability distribution of model output variables and system performance indices that results from variations and possible values of all of the input values.

Using a simple Monte Carlo analysis, values of all of the parameter sets are selected randomly from distributions describing the individual and joint uncertainty in each, and then the modelled system is simulated to obtain estimates of the selected performance indices. This must be done many times (often well over 100) to obtain a statistical description of system performance variability. The number of replications needed is generally not dependent on the number of parameters whose errors are to be analysed. One can include in the simulation the

**Figure 9.17.** Monte Carlo sampling and simulation procedure for finding distributions of output variable values based on distributions, for specified reliability levels, of input data values. This technique can be applied to one or more uncertain input variables at a time. The output distributions will reflect the combined effects of this input uncertainty over the specified ranges.



uncertainty in parameters as well as natural variability. This method can evaluate the impact of single or multiple uncertain parameters.

A significant problem that arises in such simulations is that some combinations of parameter values result in unreasonable models. For example, model performance with calibration data sets might be inconsistent with available data sets. The calibration process places interesting constraints on different sets of parameter values. Thus, such Monte Carlo experiments often contain checks that exclude combinations of parameter values that are unreasonable. In these cases the generated results are conditioned on this validity check.

Whenever sampling methods are used, one must consider possible correlations among input data values. Sampling methods can handle spatial and temporal correlations that may exist among input data values, but the existence of correlation requires defining appropriate conditional distributions.

One major limitation of applying Monte Carlo methods to estimate ranges of risk and uncertainty for model output variable values, and system performance indicator values on the basis of these output variable values, is the computing time required. To reduce the time needed to perform sensitivity analyses using sampling methods, some tricks and stratified sampling methods are available. The discussion below illustrates the idea of a simple modification (or trick) using a 'standardized' Monte Carlo analysis. The more general Latin Hypercube Sampling procedure is also discussed.

## Simple Monte Carlo Sampling

To illustrate the use of Monte Carlo sampling methods, consider again Vollenweider's empirical relationship, Equation 9.5, for the average phosphorus concentration in lakes (Vollenweider, 1976). Two hundred values of each parameter were generated independently from normal distributions with the means and variances as shown in Table 9.6.

The table contains the specified means and variances for the generated values of $L$, $q$ and $z$, and also the actual values of the means and variances of the 200 generated values of $L$, $q$, $z$ and also of the 200 corresponding generated output phosphorus concentrations, $P$. Figure 9.18 displays the distribution of the generated values of $P$.

One can see that, given the estimated levels of uncertainty, phosphorus levels could reasonably range from below 10 to above 25. The probability of generating a value greater than 20 mg/m$^3$ was 12.5%. The 5–95 percentile range was 11.1 to 23.4 mg/m$^3$. In the figure, the cumulative probability curve is rough because only 200 values of the phosphorus concentration were generated, but these are clearly enough to give a good impression of the overall impact of the errors.

## Sampling Uncertainty

In this example, the mean of the 200 generated values of the phosphorus concentration, $P$, was 17.07. However, a different set of random values would have generated a

| parameter | L | q | z | P |
|---|---|---|---|---|
| **specified means and standard deviations** | | | | |
| mean standard deviations | 680.00<br>121.21 | 10.60<br>1.67 | 84.00<br>1.82 | —<br>--- |
| **generated means and standard deviations** | | | | |
| mean standard deviations | 674.18<br>130.25 | 10.41<br>1.73 | 84.06<br>1.82 | 17.07<br>3.61 |

E020829s

**Table 9.6. Monte Carlo analysis of lake phosphorus levels.**

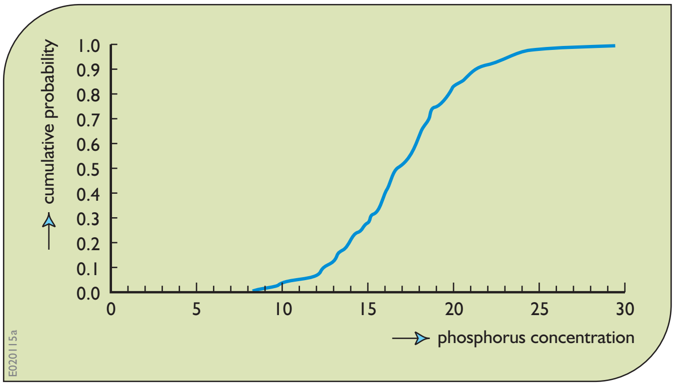



**Figure 9.18.** Distribution of lake phosphorus concentrations from Monte Carlo analysis.

| variable | $L$ | $q$ | $z$ | $P$ |
|---|---|---|---|---|
| $L$ | 1 | | | |
| $q$ | 0.079 | 1 | | |
| $z$ | 0.130 | -0.139 | 1 | |
| $P$ | 0.851 | -0.434 | 0.144 | 1 |

E020829t

**Table 9.7.** Correlation analysis of Monte Carlo results.

| | coefficient | standardized error | ratio $t$ |
|---|---|---|---|
| intercept | 18.605 | 1.790 | 10.39 |
| $L$ | 0.025 | 0.000 | 85.36 |
| $q$ | -1.068 | 0.022 | -48.54 |
| $z$ | -0.085 | 0.021 | - 4.08 |

E020829u

**Table 9.8.** Results of regression analysis on Monte Carlo Results.

different set of $P$ values as well. Thus, it is appropriate to estimate the standard error, SE, of this average. The standard error equals the standard deviation $\sigma$ of the $P$ values divided by the square root of the sample size $n$:

$$SE = \sigma/(n)^{0.5} = 3.61/(200)^{0.5} = 0.25. \tag{9.19}$$

From the central limit theorem of mathematical statistics, the average of a large number of independent values should have very nearly a normal distribution. Thus, 95% of the time, the true mean of $P$ should be in the interval $17.1 \pm 1.96(0.25)$, or 16.6 to 17.6 mg/m$^3$. This level of uncertainty reflects the observed variability of $P$ and the fact that only 200 values were generated.

## Making Sense of the Results

A significant challenge with complex models is to determine from the Monte Carlo simulation which parameter errors are important. Calculating the correlation between each generated input parameter value and the output variable value is one way of doing this. As Table 9.7 below shows, on the basis of the magnitudes of the correlation coefficients, errors in $L$ were most important, and those in $q$ were second in importance.

One can also use regression to develop a linear model defining variations in the output on the basis of errors in the various parameters. The results are shown in the Table 9.8. The fit is very good, and $R^2 = 98\%$. If the model for $P$ had been linear, a $R^2$ value of 100% should have resulted. All of the coefficients are significantly different from zero.

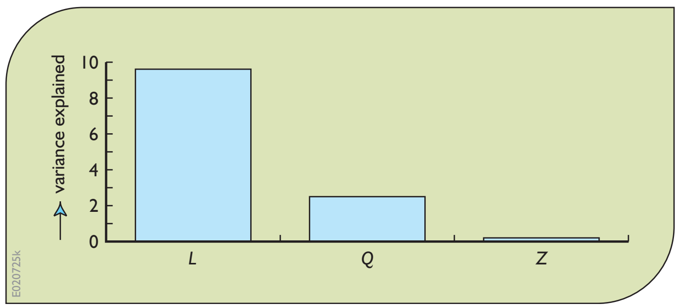



**Figure 9.19.** Reduction in the variance of $P$ that is due to dropping from the regression model each variable individually. Clearly $L$ has the biggest impact on the uncertainty in $P$, and $z$ the least.

Note that the correlation between $P$ and $z$ was positive in Table 9.7, but the regression coefficient for $z$ is negative. This occurred because there is a modest negative correlation between the generated $z$ and $q$ values. Use of partial correlation coefficients can also correct for such spurious correlations among input parameters.

Finally, we display a plot, Figure 9.19, based on this regression model, illustrating the reduction in the variance of $P$ that results from dropping each variable

| parameter | L | q | z | P |
|---|---|---|---|---|
| | specified means and standard deviations | | | |
| mean | 680.00 | 10.60 | 84.00 | — |
| standard deviations | 121.21 | 1.67 | 1.82 | --- |
| | generated means and standard deviations | | | |
| mean | 680.00 | 10.60 | 84.00 | 17.03 |
| standard deviations | 121.21 | 1.67 | 1.82 | 3.44 |

E020829v

**Table 9.9.** Standardized Monte Carlo analysis of lake phosphorus levels.

| variable | L | q | z | P |
|---|---|---|---|---|
| L | 1.00 | | | |
| q | 0.01 | 1.00 | | |
| z | 0.02 | 0.00 | 1.00 | |
| P | 0.85 | -0.50 | -0.02 | 1.00 |

E020829w

**Table 9.10.** Correlation analysis of standardized Monte Carlo results.

individually. Clearly, $L$ has the biggest impact on the uncertainty in $P$, and $z$ the least.

## Standardized Monte Carlo Analysis

Using a 'standardized' Monte Carlo analysis, one could adjust the generated values of $L$, $q$ and $z$ above so that the generated samples actually have the desired mean and variance. While making that correction, one can also shuffle their values so that the correlations among the generated values for the different parameters are near zero, as is desired. This was done for the 200 generated values to obtain the statistics shown in Table 9.9.

Repeating the correlation analysis from before (shown in Table 9.10) now yields much clearer results that are in agreement with the regression analysis. The correlation between $P$ and both $q$ and $z$ are now negative, as they should be. Because the generated values of the three parameters have been adjusted to be uncorrelated, the signal from one is not confused with the signal from another.

The mean phosphorus concentration changed very little. It is now 17.0 instead of 17.1 mg/m$^3$.

## Generalized Likelihood Estimation

Beven (1993) and Binley and Beven (1991) suggest a Generalized Likelihood Uncertainty Estimation (GLUE) technique for assessment of parameter error uncertainty using Monte Carlo simulation. It is described as a 'formal methodology for some of the subjective elements of model calibration' (Beven, 1989, p. 47). The basic idea is to begin by assigning reasonable ranges for the various parameters, and then to draw parameter sets from those ranges using a uniform or some similar (and flat) distribution. These generated parameter sets are then used on a calibration data set so that unreasonable combinations can be rejected, while reasonable values are assigned a posterior probability based upon a likelihood measure that may reflect several dimensions and characteristics of model performance.

Let $L(P_i) > 0$ be the value of the likelihood measure assigned to the calibration sequence of the $i^{\text{th}}$ parameter set. Then the model predictions generated with parameter set/combination $P_i$ are assigned posterior probability, $p(P_i)$, where

$$p(P_i) = L(P_i) \Big/ \sum_j L(P_j) \tag{9.20}$$

These probabilities reflect the form of Bayes theorem, which is well supported by probability theory (Devore, 1991). This procedure should capture reasonably well the dependence or correlation among parameters, because *reasonable* sequences will all be assigned larger probabilities, whereas sequences that are unable to reproduce the system response over the calibration period will be rejected or assigned small probabilities.

However, in a rigorous probabilistic framework, the $L$ would be the likelihood function for the calibration series for particular error distributions. (This could be checked with available goodness-of-fit procedures; for example, Kuczera, 1988.) When relatively ad hoc measures are adopted for the likelihood measure with little statistical validity, the $p(P_i)$ probabilities are best described as pseudo probabilities or 'likelihood' weights.

Another concern with this method is the potential efficiency. If the parameter ranges are too wide, a large number of unreasonable or very unlikely parameter combinations will be generated. These will either be rejected or else will have small probabilities and, thus, little effect on the analysis. In this case, the associated processing would be a waste of effort. A compromise is to use some data to calibrate the model and to generate a prior or initial distribution for the parameters that is at least centred in the best range (Beven, 1993, p. 48). Use of a different calibration period to generate the $p(P_i)$ then allows an updating of those initial probabilities to reflect the information provided by the additional calibration period with the adopted likelihood measures.

After the accepted sequences are used to generate sets of predictions, the likelihood weights would be used in the calculation of means, variances and quantiles. The resulting conditional distribution of system output reflects the initial probability distributions assigned to parameters, the rejection criteria and the likelihood measure adopted to assign 'likelihood' weights.

### Latin Hypercube Sampling

For the simple Monte Carlo simulations described in Chapters 7 and 8, with independent errors, a probability distribution is assumed for each input parameter or variable. In each simulation run, values of all input data are obtained from sampling those individual and independent distributions. The value generated for an input parameter or variable is usually independent of what that value was in any previous run, or what other input parameter or variable values are in the same run. This simple sampling approach can result in a clustering of parameter values, and hence both redundancy of information from repeated sampling in the same regions of a distribution and lack of information because there is no sampling in other regions of the distributions.

A stratified sampling approach ensures more even coverage of the range of input parameter or variable values with the same number of simulation runs. This can be accomplished by dividing the input parameter or variable space into sections and sampling from each section with the appropriate probability.

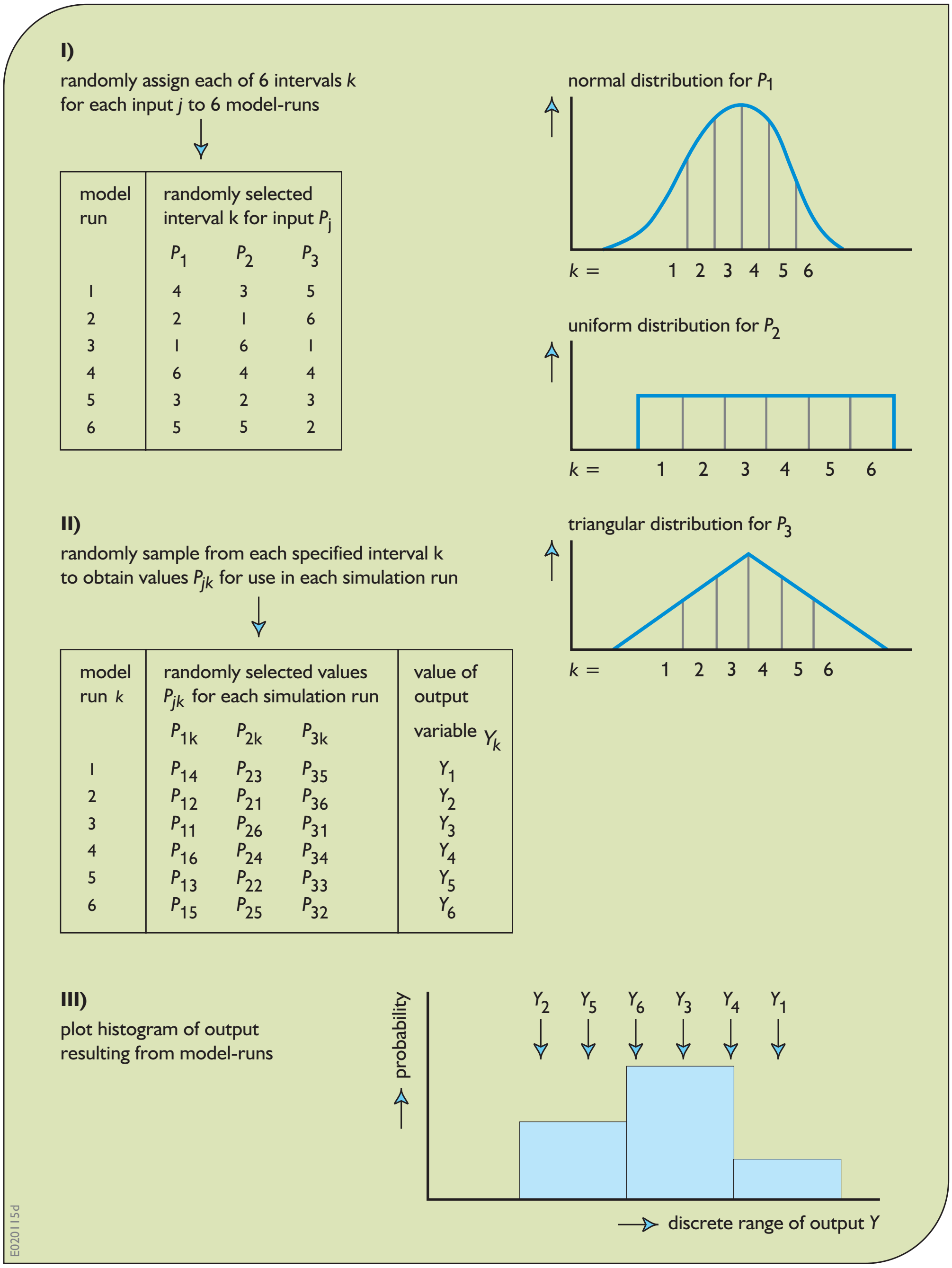
One such approach, Latin hypercube sampling (LHS), divides each input distribution into sections of equal probability for the specified probability distribution, and draws one observation randomly from each section. Hence the ranges of input values within each section actually occur with equal frequency in the experiment. These values from each section for each distribution are randomly assigned to those from other sections to construct sets of input values for the simulation analysis. Figure 9.20 shows the steps in constructing an LHS for six simulations involving three inputs $P_j$ ($P_1$, $P_2$, and $P_3$) and six intervals of their respective normal, uniform and triangular probability distributions.

# 5. Performance Indicator Uncertainties

## 5.1. Performance Measure Target Uncertainty

Another possible source of uncertainty is the selection of performance measure target values. For example, consider a target value for a pollutant concentration based on the effect of exceeding it in an ecosystem. Which target value is best or correct? When this is not clear, there are various ways of expressing the uncertainty associated with any target value. One such method is the use of fuzzy sets (Chapter 5). Use of 'grey' numbers or intervals instead of 'white' or fixed target values is another. When some uncertainty or disagreement exists over the selection of the best target value for a particular performance measure, it seems to us the most direct and transparent way to do this is to subjectively assign a distribution over a range of possible target values. Then this subjective probability distribution can be factored into the tradeoff analysis, as outlined in Figure 9.21.

One of the challenges associated with defining and including in an analysis the uncertainty associated with a target or threshold value for a performance measure is that of communicating just what the result of such an analysis means. Referring to Figure 9.21, suppose the target value represents some maximum limit of a pollutant concentration, such as phosphorus, in the flow during a given period of time at a given site or region, and it is not certain just what that maximum limit should be. Subjectively defining the distribution of that maximum limit, and considering that uncertainty along with the




| model run | randomly selected interval k for input $P_j$ | | |
|---|---|---|---|
| | $P_1$ | $P_2$ | $P_3$ |
| 1 | 4 | 3 | 5 |
| 2 | 2 | 1 | 6 |
| 3 | 1 | 6 | 1 |
| 4 | 6 | 4 | 4 |
| 5 | 3 | 2 | 3 |
| 6 | 5 | 5 | 2 |

| model run k | randomly selected values $P_{jk}$ for each simulation run | | | value of output variable $Y_k$ |
|---|---|---|---|---|
| | $P_{1k}$ | $P_{2k}$ | $P_{3k}$ | |
| 1 | $P_{14}$ | $P_{23}$ | $P_{35}$ | $Y_1$ |
| 2 | $P_{12}$ | $P_{21}$ | $P_{36}$ | $Y_2$ |
| 3 | $P_{11}$ | $P_{26}$ | $P_{31}$ | $Y_3$ |
| 4 | $P_{16}$ | $P_{24}$ | $P_{34}$ | $Y_4$ |
| 5 | $P_{13}$ | $P_{22}$ | $P_{33}$ | $Y_5$ |
| 6 | $P_{15}$ | $P_{25}$ | $P_{32}$ | $Y_6$ |

**Figure 9.20.** Schematic representation of a Latin hypercube sampling procedure for six simulation runs.

uncertainty (probability of exceedance function) of pollutant concentrations – the performance measure – one can attach a reliability to any probability of exceeding the maximum desired concentration value.

The 95% probability of exceedance shown on Figure 9.21, say $P_{0.95}$, should be interpreted as: 'we can be 95% confident that the probability of the maximum desired pollutant concentration being exceeded will be no greater than $P_{0.95}$.' We can be only 5% confident that the probability of exceeding the desired maximum concentration will be no greater than the lower $P_{0.05}$ value. Depending on whether the middle line through the subjective distribution of target values in Figure 9.21 represents the most likely or median target value, the associated probability of

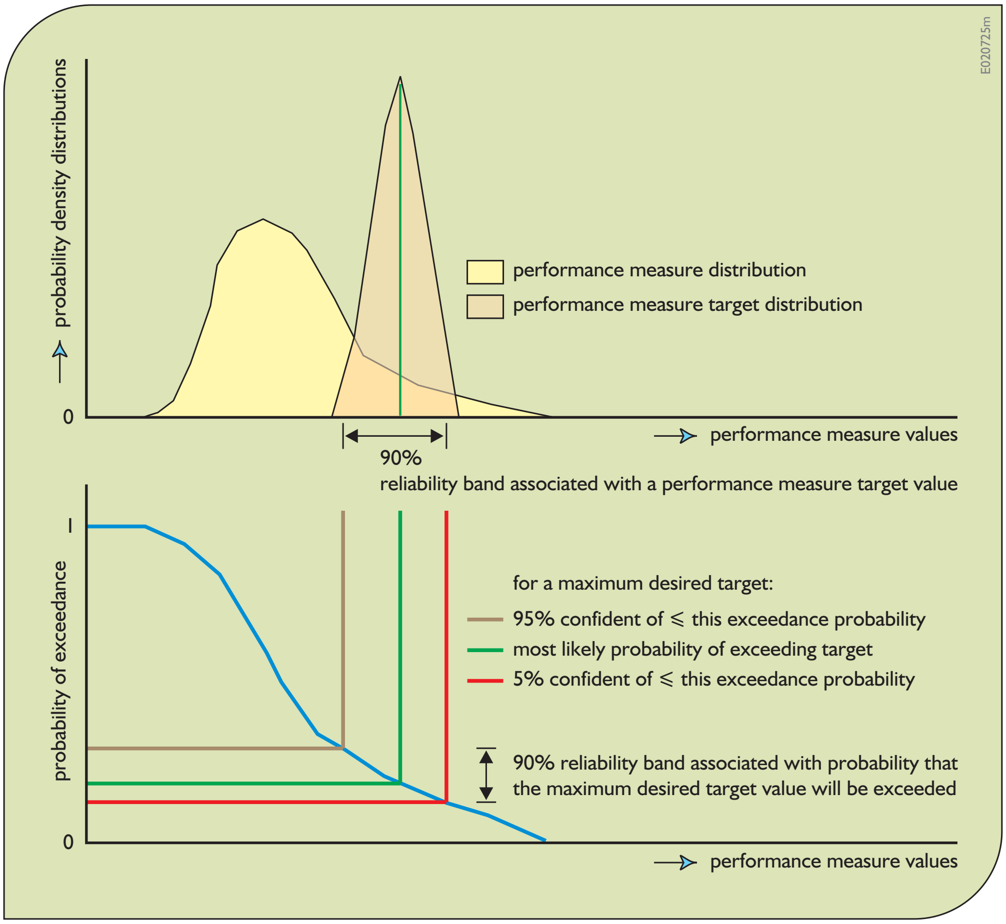



**Figure 9.21.** Combining the probability distribution of performance measure values with the probability distribution of performance measure target values to estimate the confidence one has in the probability of exceeding a maximum desired target value.

exceedance is either the most likely, as indicated in Figure 9.21, or that for which we are only 50% confident.

Figure 9.22 attempts to show how to interpret the reliabilities when the uncertain performance targets are:

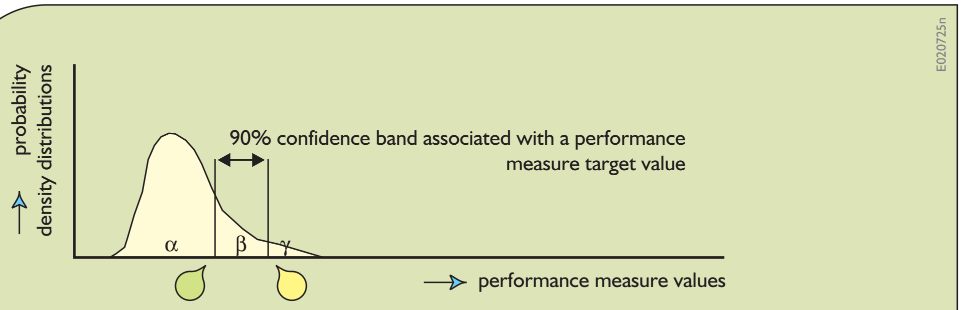
- minimum acceptable levels that are to be maximized,
- maximum acceptable levels that are to be minimized, or
- optimum levels.

An example of a minimum acceptable target level might be the population of wading birds in an area. An example of a maximum acceptable target level might be, again, the phosphorus concentration of the flow in a specific wetland or lake. An example of an optimum target level might be the depth of water most suitable for selected species of aquatic vegetation during a particular period of the year.

For performance measure targets that are not expressed as minimum or maximum limits but that are the 'best' values, referring to Figure 9.22, one can state that there is a 90% reliability that the probability of achieving the desired target is no more than $\beta$. The 90% reliability level of not achieving the desired target is at least $\alpha + \gamma$. The probability of the performance measure being too low is at least $\alpha$, and the probability of the performance measure being too high is at least $\gamma$, again at the 90% reliability levels. As the reliability level decreases, the bandwidth decreases, and the probability of not meeting the target increases.

Now, clearly there is uncertainty associated with each of these uncertainty estimations, and this raises the question of how valuable is the quantification of the uncertainty of each additional component of the plan in an evaluation process. Will plan-evaluators and decision-makers benefit from this additional information, and just how much additional uncertainty information is useful?

Now consider again the tradeoffs that need to be made, as illustrated in Figure 9.7. Instead of considering a single target value, as shown on the figure, assume there is a 90% reliability or probability range associated with that single performance measure target value. Also




**For a performance measure that is to be maximized:**

5% confident that the probability of meeting or exceeding the desired target is at least β+γ.

5% confident that the probability of not meeting or exceeding the desired target is less than α.

**For a performance measure that is to be maximized:**

95% confident that the probability of meeting or exceeding the desired target is at least γ.

95% confident that the probability of not meeting or exceeding the desired target is less than α+β.

**For a performance measure that is to be minimized:**

95% confident that the probability of exceeding a desired target is no greater than β+γ.

95% confident that the probability of not exceeding the desired target is at least α.

**For a performance measure that is to be minimized:**

5% confident that the probability of exceeding a desired target is no greater than γ.

5% confident that the probability of not exceeding the desired target is at least α+β.

**Figure 9.22.** Interpreting the results of combining performance measure probabilities with performance measure target probabilities depends on the type of performance measure. The letters $\alpha$, $\beta$ and $\gamma$ represent proportions of the probability density function of performance measure values. (Hence probabilities $\alpha + \beta + \gamma = 1$.)

assume that the target is a maximum desired upper limit, e.g. of some pollutant concentration.

In the case shown in Figure 9.23, the tradeoff is clearly between cost and reliability. In this example, no matter what reliability one chooses, Plan A is preferred to Plan B with respect to reliability, but Plan B is preferred to Plan A with respect to cost. The tradeoff is only between these two performance indicators or measures.

Consider, however, a third plan, as shown in Figure 9.24. This situation adds to the complexity of making appropriate tradeoffs. Now there are three criteria: cost, probability of exceedance and the reliabilities of those probabilities. Add to this the fact that there will be multiple performance measure targets, each expressed in terms of their maximum probabilities of exceedance and the reliability of those probabilities.

In Figure 9.24, in terms of cost the plans are ranked, from best to worst, B, C and A. In terms of the 90 reliable probability, they are ranked A, B and C, but at the 50 percent reliability level the ranking is A, C and B.

If the plan evaluation process has difficulty handling all this, it may indicate the need to focus the uncertainty analysis effort on just what is deemed important, achievable and beneficial. Then, when the number of alternatives has been narrowed down to only a few that appear to be the better ones, a more complete uncertainty analysis can be performed. There is no need and no benefit in performing sensitivity and uncertainty analyses on all possible management alternatives. Rather one can focus on those alternatives that look most promising, and then carry out additional uncertainty and sensitivity analyses only when important uncertain performance indicator values demands more scrutiny.

## 5.2. Distinguishing Differences Between Performance Indicator Distributions

Simulations of alternative water management infrastructure designs and operating policies require a comparison of the simulation outputs – the performance measures or indicators – associated with each alternative. A reasonable

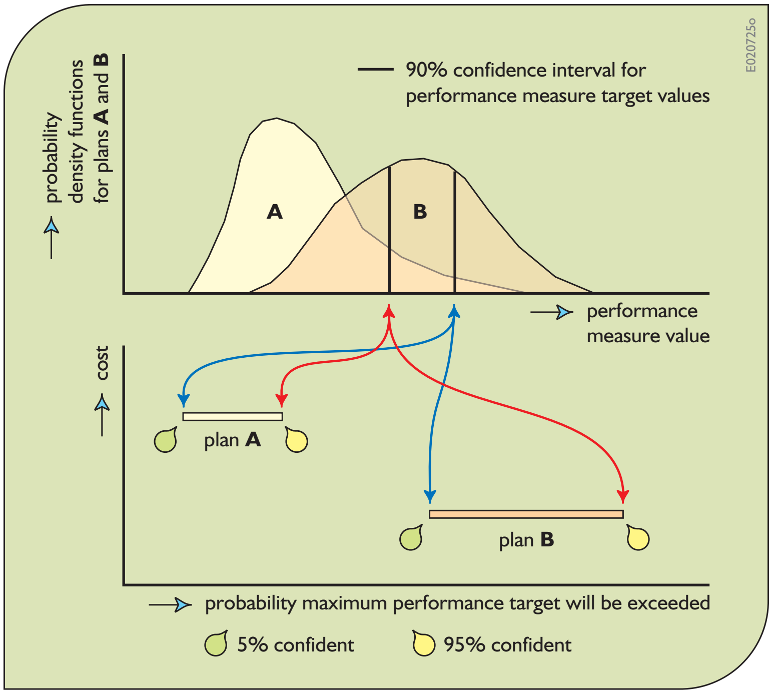



**Figure 9.23.** Two plans showing ranges of probabilities, depending on the reliability, that an uncertain desired maximum (upper limit) performance target value will be exceeded. The 95% reliability levels are associated with the higher probabilities of exceeding the desired maximum target. The 5% reliability levels are associated with the more desirable lower probabilities of exceeding the desired maximum target. Plan A with reduced probabilities of exceeding the upper limit costs more than Plan B.

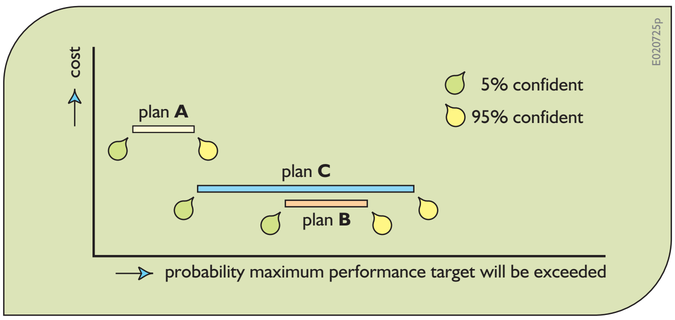



**Figure 9.24.** Tradeoffs among cost, probabilities and the reliability or confidence level of those probabilities. The relative ranking of plans with respect to the probability of exceeding the desired (maximum limit) target may depend on the reliability or confidence associated with that probability.

question to ask is whether the observed differences are statistically significant. Can one really tell whether one alternative is better than another, or whether the observed differences are explainable by random variations attributable to variations in the inputs and the way the system responds?

This is a common statistical issue that is addressed by standard hypothesis tests (Devore, 1991; Benjamin and Cornell, 1970). Selection of an appropriate test requires that one first resolve what type of change one expects in the variables. To illustrate, consider the comparison of two different operating policies. Let $Y_1$ denote the set of output performance variable values with the first policy, and $Y_2$ the set of output performance variable values of the second policy. In many cases, one would expect one policy to be better than the other. One measure might be the difference in the mean of the variables; for example, $E[Y_1] < E[Y_2]$. Alternatively one could check the difference in the median ($50^{th}$ percentile) of the two distributions.

In addition, one could look for a change in the variability or variance, or a shift in both the mean and the variance. Changes described by a difference in the mean or median often make the most sense, and many statistical tests are available that are sensitive to such changes. For such investigations parametric and non-parametric tests for paired and unpaired data can be employed.

Consider the differences between 'paired' and 'unpaired' data. Suppose that the meteorological data for 1941 to 1990 is used to drive a simulation model generating data as described in Table 9.11.

Here there is one sample, $Y_1(1)$ through $Y_1(50)$, for Policy 1, and another sample, $Y_2(1)$ through $Y_2(50)$, for Policy 2. However, the two sets of observations are not independent. For example, if 1943 was a very dry year,

| | | |
|---|---|---|
| 1941 | $Y_1$ ( 1) | $Y_2$ ( 1) |
| 1942 | $Y_1$ ( 2) | $Y_2$ ( 2) |
| 1943 | $Y_1$ ( 3) | $Y_2$ ( 3) |
| 1944 | $Y_1$ ( 4) | $Y_2$ ( 4) |
| 1989 | $Y_1$ (49) | $Y_2$ (49) |
| 1990 | $Y_1$ (50) | $Y_2$ (50) |

E020829x

**Table 9.11.** Possible flow data from a fifty-year simulation.

then we would expect both $Y_1(3)$ for Policy 1 in that year and $Y_2(3)$ for Policy 2 to be unusually small. With such paired data, one can use a paired hypothesis test to check for differences. Paired tests are usually easier than the corresponding unpaired tests that are appropriate in other cases. For example, if one were checking for a difference in average rainfall depth between the periods 1941 to 1960 and 1961 to 1990, then one would have two sets of independent measurements for the two periods. With such data, one should use a two-sample unpaired test.

Paired tests are generally based on the differences between the two sets of output, $Y_1(i) - Y_2(i)$. These are viewed as a single independent sample. The question is then whether the differences are positive (say $Y_1$ tends to be larger then $Y_2$) or negative ($Y_1$ tends to be smaller), or whether positive and negative differences are equally likely (there is no difference between $Y_1$ and $Y_2$).

Both parametric and non-parametric families of statistical tests are available for paired data. The common parametric test for paired data (a one-sample $t$ test) assumes that the mean of the differences

$$X(i) = Y_1(i) - Y_2(i) \tag{9.21}$$

are normally distributed. The hypothesis of no difference is rejected if the t statistic is sufficiently large, given the sample size $n$.

Alternatively, one can employ a non-parametric test and avoid the assumption that the differences $X$(i) are normally distributed. In such a case, one can use the Wilcoxon Signed Rank test. This non-parametric test ranks the absolute values $|X(i)|$ of the differences. If the sum $S$ of the ranks of the positive differences deviates sufficiently from its expected value, $n(n + 1)/4$ (were there no difference between the two distributions), one can conclude that there is a statistically significant difference between the $Y_1(i)$ and $Y_2(i)$ series. Standard statistical texts have tables of the distribution of the sum $S$ as a function of the sample size $n$, and provide a good analytical approximation for $n > 20$ (for example, Devore, 1991). Both the parametric $t$ test and the non-parametric Wilcoxon Signed Rank test require that the differences between the simulated values for each year be computed.

## 6. Communicating Model Output Uncertainty

Spending money on reducing uncertainty would seem preferable to spending it on ways of calculating and describing it better. Yet attention to uncertainty communication is critically important if uncertainty analyses and characterizations are to be of value in a decision-making process. In spite of considerable efforts by those involved in risk assessment and management, we know very little about how to ensure effective risk communication to gain the confidence of stakeholders, incorporate their views and knowledge, and influence favourably the acceptability of risk assessments and risk-management decisions.

The best way to communicate concepts of uncertainty may well depend on what the audiences already know about risk and the various types of probability distributions (e.g. density, cumulative, exceedance) based on objective and subjective data, and the distinction between mean or average values and the most likely values. Undoubtedly graphical representations of these ways of describing uncertainty considerably facilitate communication.

The National Research Council (NRC, 1994) addressed the extensive uncertainty and variability associated with estimating risk and concluded that risk characterizations should not be reduced to a single number or even to a range of numbers intended to portray uncertainty. Instead, the report recommended that managers and the interested public should be given risk characterizations that are both qualitative and quantitative, and both verbal and mathematical.

In some cases, communicating qualitative information about uncertainty to stakeholders and the public in general may be more effective than providing quantitative information. There are, of course, situations in which quantitative uncertainty analyses are likely to provide information that is useful in a decision-making process. How else can tradeoffs such as those illustrated in Figures 9.10 and 9.27 be identified? Quantitative uncertainty analysis can often be used as the basis of qualitative information about uncertainty, even if the quantitative information is not what is communicated to the public.

One should acknowledge to the public the widespread confusion regarding the differences between variability and uncertainty. Variability does not change through further measurement or study, although better sampling can improve our knowledge about it. Uncertainty reflects gaps in information about scientifically observable phenomena.

While it is important to communicate prediction uncertainties and the reliabilities of those uncertainties, it is equally important to clarify who or what is at risk, the possible consequences, and the severity and irreversibility of an adverse effect should a target value, for example, not be met. This qualitative information is often critical to informed decision-making. Risk and uncertainty communication is always complicated by the reliability and amounts of available relevant information as well as how that information is presented. Effective communication between people receiving information about who or what is at risk, or what might happen and just how severe and irreversible an adverse effect might be should a target value not be met, is just as important as the level of uncertainty and the reliability associated with such predictions. A two-way dialogue between those receiving such information and those providing it can help identify just what seems best for a particular audience.

Risk and uncertainty communication is a two-way activity that involves learning and teaching. Communicators dealing with uncertainty should learn about the concerns and values of their audience, their relevant knowledge and their experience with uncertainty issues. Stakeholders' knowledge of the sources and reasons for uncertainty needs to be incorporated into assessment and management and communication decisions. By listening, communicators can craft risk messages that better reflect the perspectives, technical knowledge and concerns of the audience.

Effective communication should begin before important decisions have been made. It can be facilitated in communities by citizen advisory panels, which can give planners and decision-makers a better understanding of the questions and concerns of the community, and an opportunity to test their effectiveness in communicating concepts and specific issues regarding uncertainty.

One approach to make risk more meaningful is to make risk comparisons. For example, a ten parts-per-billion target for a particular pollutant concentration is equivalent to ten seconds in over thirty-one years. If this is an average daily concentration target that is to be satisfied '99%' of the time, then this is equivalent to an expected violation of less than one day every three months.

Many perceive the reduction of risk by an order of magnitude as though it were a linear reduction. A better way to illustrate orders of magnitude of risk reduction is shown in Figure 9.25, in which a bar graph depicts better than words that a reduction in risk from one in a 1,000 ($10^{-3}$) to one in 10,000 ($10^{-4}$) is a reduction of 90% and that a further reduction to one in 100,000 ($10^{-5}$) is a reduction tenfold less than the first reduction of 90%. The percent of the risk that is reduced by whatever measures is a much easier concept to communicate than reductions expressed in terms of estimated absolute risk levels, such as $10^{-5}$.

Risk comparisons can be helpful, but they should be used cautiously and tested if possible. There are dangers in comparing risks of diverse character, especially when the intent of the comparison is seen as reducing a risk (NRC, 1989). One difficulty in using risk comparisons is that it is not always easy to find risks that are sufficiently similar to make a comparison meaningful. How is one to compare two alternatives that have two different costs and two different risk levels, for example, as shown in Figure 9.23? One way is to perform an indifference analysis (Chapter 10), but that can lead to differing results depending who performs it. Another way is to develop utility functions using weights, where, for example reduced phosphorus load by half is equivalent to a 25% shorter hydroperiod in that area, but again each person's utility or tradeoff may differ.

At a minimum, graphical displays of uncertainty can be helpful. Consider the common system performance indicators that include:

- time-series plots for continuous time-dependent indicators (Figure 9.26 upper left),
- probability exceedance distributions for continuous indicators (Figure 9.26 upper right),
- histograms for discrete event indicators (Figure 9.26 lower left),
- overlays on maps for space-dependent discrete events (Figure 9.26 lower right).

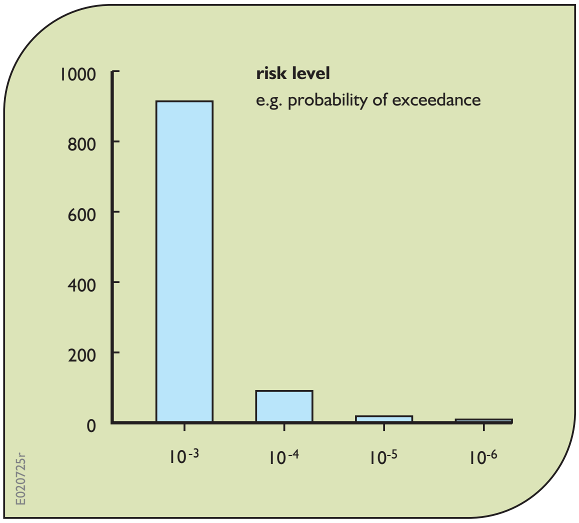



**Figure 9.25.** Reducing risk (x-axis) by orders of magnitude is not equivalent to linear reductions in some risk indicator (y-axis).

The first three graphs in Figure 9.26 could show, in addition to the single curve or bar that represents the most likely output, a range of outcomes associated with a given confidence level. For overlays of information on maps, different colours could represent the spatial extents of events associated with different ranges of risk or uncertainty. Figure 9.27, corresponding to Figure 9.26, illustrates these approaches for displaying these ranges.

## 7. Conclusions

This chapter has provided an overview of uncertainty and sensitivity analyses in the context of hydrological or water resources systems simulation modelling. A broad range of tools are available to explore, display and quantify the sensitivity and uncertainty in predictions of key output variables and system performance indices with respect to imprecise and random model inputs and to assumptions concerning model structure. They range from relatively

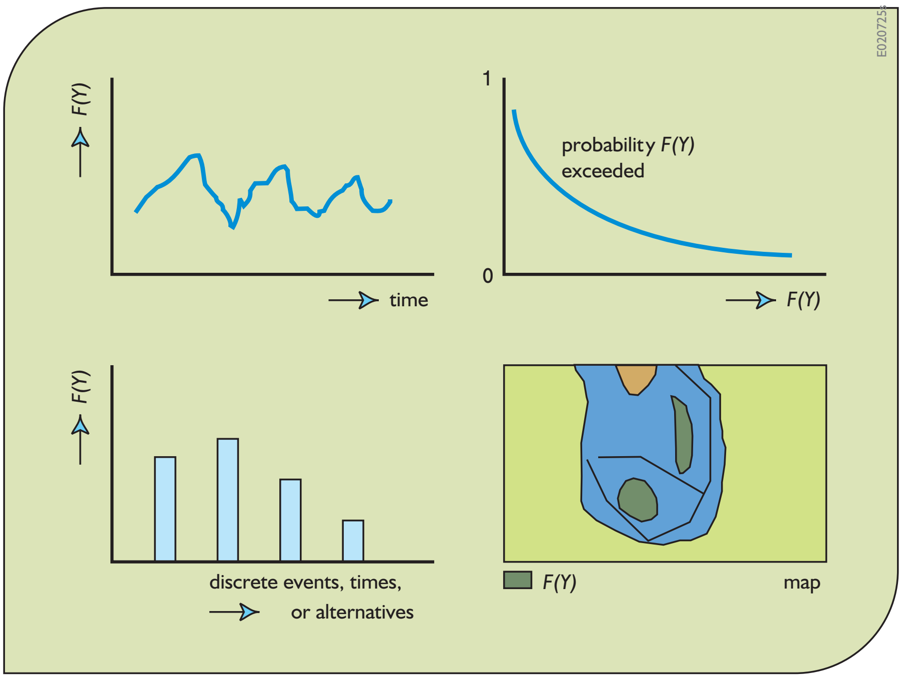



**Figure 9.26.** Different types of displays used to show model output *Y* or system performance indicator values *F*(*Y*).

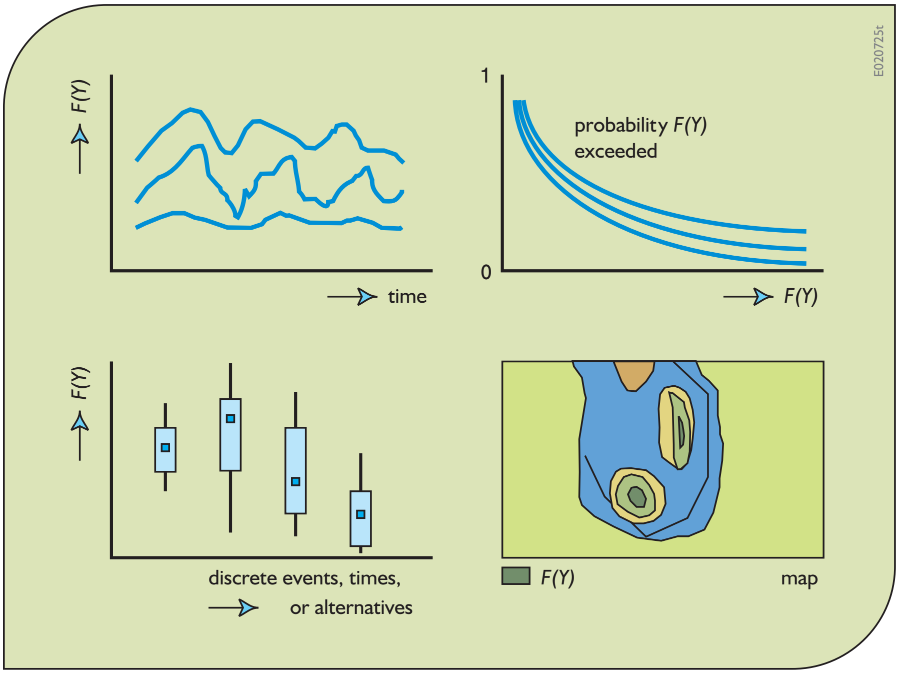



**Figure 9.27.** Plots of ranges of possible model output $Y$ or system indicator values $F(Y)$ for different types of displays.

simple deterministic sensitivity analysis methods to more involved first-order analyses and Monte Carlo sampling methods.

Because of the complexity of many watersheds or river basins, using Monte Carlo methods for uncertainty analyses may be a very large and unattractive undertaking. It is therefore often prudent to begin with the relatively simple deterministic procedures. These, coupled with a probabilistically based first-order uncertainty analysis method, can help quantify the uncertainty in key output variables and system performance indices, and the relative contributions of uncertainty in different input variables to the uncertainty in different output variables and system performance indices. These relative contributions may differ depending upon which output variables and indices are of interest.

A sensitivity analysis can provide a systematic assessment of the impact of parameter value imprecision on output variable values and performance indices, and of the relative contribution of errors in different parameter values to that output uncertainty. Once the key variables are identified, it should be possible to determine the extent to which parameter value uncertainty can be reduced through field investigations, development of better models and other efforts.

Model calibration procedures can be applied to individual catchments and subsystems, as well as to composite systems. Automated calibration procedures have several advantages, including the explicit use of an appropriate statistical objective function, identification of those parameters that best reproduce the calibration data set with the given objective function, and the estimations of the statistical precision of the estimated parameters.

All of these tasks together can represent a formidable effort. However, knowledge of the uncertainty associated with model predictions can be as important to management decisions and policy formulation as are the predictions themselves.

No matter how much attention is given to quantifying and reducing uncertainties in model outputs, uncertainties will remain. Professionals who analyse risk, managers and decision-makers who must manage risk, and the public who must live with risk and uncertainty, have different information needs and attitudes regarding risk and uncertainty. It is clear that information needs differ among those who model or use models, those who make substantial investment or social decisions, and those who are likely to be affected by those decisions. Meeting those needs should result in more informed decision-making, but this comes at a cost that should be considered along with the benefits of having this sensitivity and uncertainty information.

## 8. References

BECK, M.B. 1987. Water quality modelling: a review of the analysis of uncertainty. *Water Resources Research*, Vol. 23, No. 8, pp. 1393–1442.

BENAMAN, J. 2002. *A systematic approach to uncertainty analysis for a distributed watershed model*. Ph.D. Thesis, School of Civil and Environmental Engineering, Cornell University, Ithaca, N.Y.

BENJAMIN, J.R. and CORNELL, C.A. 1970. *Probability, statistics and decision for civil engineers*. New York, McGraw-Hill.

BEVEN, K. 1989. Changing ideas in hydrology: the case of physically-based models. *Journal of Hydrology*. No. 105, pp. 157–72.

BEVEN, K. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, No. 16, 41–51.

BINLEY, A.M. and BEVEN, K.J. 1991. Physically-based modelling of catchment hydrology: a likelihood approach to reducing predictive uncertainty. In: D.G. Farmer and M.J. Rycroft (eds.), *Computer modelling in the environmental sciences*, Oxford, Clarendon, pp. 75–88.

DEGARMO, E.P.; SULLIVAN, W.G. and BONTADELLI, J.A. 1993. *Engineering economy*. New York, MacMillan.

DEVORE, J. 1991. *Probability and statistics for engineering and the sciences*, 3rd Edition. London, Brooks/Cole.

ESCHENBACH, T.G. 1992. Spider plots versus tornado diagrams for sensitivity analysis. *Interfaces*, No. 22, pp. 40–6.

GAVEN, D.C. and BURGES, S.J. 1981. Approximate error bounds for simulated hydrograph. *Journal of Hydraulic Engineering*, Vol. 107, No. 11, pp. 1519–34.

KUCZERA, G. 1988. On the validity of first-order prediction limits for conceptual hydrological models. *Journal of Hydrology*, No. 103, pp. 229–47.

LAL, W. 1995. *Sensitivity and uncertainty analysis of a regional model for the natural system of South Florida*. West Palm Beach, Fla., South Florida Water Management District. Draft report, November.

MCCUEN, R. 1973. The role of sensitivity analysis in hydrological modeling. *Journal of Hydrology*, Vol. 18, No. 1, pp. 37–53.

NRC (National Research Council). 1989. *Improving risk communication*. Washington, D.C., National Academy Press.

NRC (National Research Council). 1994. *Science and judgment in risk assessment*. Washington, D.C., National Academy Press.

STEDINGER, J.R. and TAYLOR, M.R. 1982. Synthetic streamflow generation, part II: Effect of parameter uncertainty. *Water Resources Research*, Vol. 18, No. 4, pp. 919–24.

VOLLENWEIDER, R.A. 1976. Advances in defining critical loading levels for phosphorus in lake eutrophication, *Memorie dell'Istituto Italiano di Idrobiologia*, No. 33, pp. 53–83.

### Additional References (Further Reading)

ANDERSON, J.L. 1998. Embracing uncertainty: the interface of Bayesian statistics and cognitive psychology. *Conservation Ecology* (online), Vol. 2, No. 1, p. 2. Available from the Internet. URL: http://www.consecol.org/vol2/iss1/art2

BERRY, D.A. 1996. *Statistics: a Bayesian perspective*. Belmont, Calif., Duxbury.

CHAPRA, S.C. and RECKHOW, K.H. 1979. Expressing the phosphorus loading concept in probabilistic terms. *Journal of the Fisheries Research Board of Canada*, No. 36, pp. 225–9.

COLWELL, R.K. 1974. Predictability, consistency, and contingency of periodic phenomena. *Ecology*, Vol. 55, pp. 1148–53.

COVELLO, V.T. and MERKHOFER, M.W. 1993. *Risk assessment methods*. London, Plenum.

COVELLO, V.T. 1987. Decision analysis and risk management decision-making: issues and methods. *Risk Analysis*, No. 7, pp. 131–9.

DEUTSCH, C.V. and JOURNEL, A.G. 1992. *GS-LIB: geostatistical software library and user's guide*. New York, Oxford University Press.

DEVORE, J.L. and PECK, R. 1996. *Statistics: the exploration and analysis of data*, 3rd edn. London, Brooks/Cole.

DILKS, D.W.; CANALE, R.P. and MEIER, P.G. 1992. Development of Bayesian Monte Carlo techniques for water quality model uncertainty. *Ecological Modelling*, Vol. 62, pp. 149–62.

DILKS, D.W. and JAMES, R.T. 2002. *Application of Bayesian Monte Carlo analysis to determine the uncertainty in the Lake Okeechobee water quality model*. Proceedings of Watershed 2002. Alexandria, Va., Water Environment Federation.

DOE (Department of Energy). 1998. *Screening assessment and requirements for a comprehensive assessment*. DOE/RL-96-16, Rev. 1. Richland, Washington, US Department of Energy.

DUAN, Q.; SOROOSIAN, S. and IBBITT, R.P. 1988. A maximum likelihood criterion for use with data collected at unequal time intervals. *Water Resources Research*, Vol. 24, No. 7, pp. 1163–73.

FEDRA, K.A. 1983. Monte Carlo approach to estimation and prediction. In: M.B. Beck and G. Van Straten (eds.), *Uncertainty and forecasting of water quality*, Berlin, Springer-Verlag, pp. 259–91.

FITZ, H.C.; VOINOV, A. and COSTANZA, R. 1995. *The Everglades landscape model: multiscale sensitivity analysis*. West Palm Beach, Fla., South Florida Water Management District, Everglades Systems Research Division.

FONTAINE, T.A. and JACOMINO, V.M.F. 1997. Sensitivity analysis of simulated contaminated sediment transport. *Journal of the American Water Resources Association*, Vol. 33, No. 2, pp. 313–26.

FREY, H.C. and PATIL, S.R. 2002. Identification and review of sensitivity analysis methods. *Risk Analysis*, Vol. 22, No. 3, pp. 553–78.

GARDNER, R.H.; ROJDER, B. and BERGSTROM, U. 1983. *PRISM: a systematic method for determining the effect of parameter uncertainties on model predictions*. Technical Report, Studsvik Energiteknik AB report NW-83-/555. Sweden, Nykoping.

GELMAN, A.; CARLIN, J.; STERN, H. and RUBIN, D.B. 1995. *Bayesian data analysis*. London, Chapman and Hall.

GLEICK, J. 1987. *Chaos: making a new science*. New York, Penguin.

GUPTA, V.K. and SOROOSHIAN, S. 1985a. The automatic calibration of conceptual watershed models using derivative-based optimization algorithms. *Water Resources Research*, Vol. 21, No. 4, pp. 473–85.

GUPTA, V.K. and SOROOSHIAN, S. 1985b. The relationship between data and the precision of parameter estimates of hydrological models. *Journal of Hydrology*, No. 81, pp. 57–7.

HAIMES, Y.Y. 1998. *Risk modelling, assessment and management*. New York, Wiley.

HARLIN, J. and KUNG, C.-S. 1992. Parameter Uncertainty in simulation of design of floods in Sweden. *Journal of Hydrology*, No. 137, pp. 2009–230.

HENDRICKSON, J.D.; SOROOSHIAN, S. and BRAZIL, L.E. 1988. Comparison of Newton-type and direct-search algorithms for calibration of conceptual rainfall–runoff models. *Water Resources Research*, Vol. 24, No. 5, pp. 691–700.

HOLLING, C.S. 1978. *Adaptive environmental assessment and management*. Chichester, UK, Wiley.

IBREKK, H. and MORGAN, M.G. 1987. Graphical communication of uncertain quantities to nontechnical people. *Risk Analysis*, Vol. 7, No. 4, pp. 519–29.

ISAAKS, E.H. and SRIVASTAVA, R.M. 1989. *An introduction to applied geostatistics*. New York, Oxford University Press.

JAFFE, P.R.; PANICONI, C. and WOOD, E.F. 1988. Model calibration based on random environmental fluctuations. *Journal of Environmental Engineering*, ASCE, Vol. 114, No. 5, pp. 1136–45.

JAIN, S. and LALL, U. 2001. Floods in a changing climate: does the past represent the future? *Water Resources Research*, Vol. 37, No. 12, pp. 3193–205 DEC.

JENSEN, F.V. 2001. *Bayesian networks and decision graphs*. New York: Springer.

KANN, A. and WEYANT, J.P. 1999. A comparison of approaches for performing uncertainty analysis in integrated assessment models. *Journal of Environmental Management and Assessment*, Vol. 5, No. 1, pp. 29–46.

KELLY, E. and CAMPBELL, K. 2000. Separating variability and uncertainty: making choices. *Human and Ecological Risk Assessment, An International Journal*, Vol. 6, No. 1, February, pp. 1–13.

KELLY, E. and ROY-HARRISON, W. 1998. A mathematical construct for ecological risk: a useful framework for assessments. *Human and Ecological Risk Assessment, An International Journal*, Vol. 4, No. 2, pp. 229–41.

KELLY, E.J.; CAMPBELL, K. and HENRION, M. 1997. To separate or not to separate – that is the question: a discourse on separating variability and uncertainty in environmental risk assessments. *Learned Discourses, Society of Environmental Toxicology and Chemistry (SETAC) News*, November.

LAL, A.M.W. 2000. Numerical errors in groundwater and overland flow models. *Water Resources Research*, Vol. 36, No. 5, May, pp. 1237–48.

LAL, W.; OBEYSEKERA, J. and VAN ZEE, R. 1997. Sensitivity and uncertainty analysis of a regional simulation model for the natural system in South Florida. In: *Managing Water: Coping with Scarcity and Abundance*. Proceedings of the 27th Congress of International Association for Hydraulic Research, San Francisco, Calif., August.

LEMONS, J. (ed.). 1996. *Scientific uncertainty and environmental problem solving*. Cambridge, Mass. Blackwell.

LOPEZ, A. and LOUCKS, D.P. 1999. *Uncertainty representations in water quality prediction*. Manuscript, Civil and Environmental Engineering, Cornell University, Ithaca, N.Y.

LOUCKS, D.P.; STEDINGER, J.R. and HAITH, D.A. 1981. *Water resource systems planning and analysis*. Englewood Cliffs, N.J., Prentice-Hall.

LUDWIG, D.; HILBORN, R. and WALTERS, C. 1993. Uncertainty, resource exploitation, and conservation: lessons from history. *Science*, Vol. 260, pp. 17–36.

MAJONI, H. and QUADE, E.S. 1980. Pitfalls of analysis. New York, Wiley.

MCCARTHY, J.J.; CANZIANI, O.F.; LEARY, N.A.; DOKKEN, D.J. and WHITE, K.S. (eds.). 2001. Climate change 2001: impacts, adaptation, and vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change, http://www.ipcc.ch/ pub/tar/wg2/index.htm

MCCUEN, R.H. and SNYDER, W.M. 1983. *Hydrological Modelling: statistical methods and applications*. Englewood Cliffs, N.J., Prentice-Hall.

MEIXNER, T.; GUPTA, H.V.; BASTIDAS, L.A. and BALES, R.C. 1999. Sensitivity analysis using mass flux and concentration. *Hydrological Processes*, Vol. 13, pp. 2233–44.

MISER, H.J. 1980. Operations research and systems analysis. *Science*, No. 209, pp. 174–82.

MORGAN, M.G. and HENRION, M. 1990. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge, UK, Cambridge University Press.

NEARING, M.A.; DEER-ASCOUGH, L. and LAFLEN, J.F. 1990. Sensitivity analysis of the WEPP hillslope profile erosion model. *Transactions of the American Society of Agricultural Engineers*, Vol. 33, No. 3, pp. 839–49.

NRC (National Research Council). 1996. *Committee on risk characterization, understanding risk: informing decision in a democratic society*. In: P.S. Stern and H.V. Fineberg (eds.), Washington, D.C., National Academy Press.

PHILLIPS, D.L. and MARKS, D.G. 1996. Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. *Ecological Modelling*, No. 91, pp. 213–29.

PRESS, S.J. and TANUR, J.M. 2001. *The subjectivity of scientists and the Bayesian approach*. New York: Wiley.

RECKHOW, K.H. 1994. Water quality simulation modelling and uncertainty analysis for risk assessment and decision-making. *Ecological Modelling*, No. 72, pp. 1–20.

RECKHOW, K.H. 1999. Water quality prediction and probability network models. *Canadian Journal of Fisheries and Aquatic Sciences*, No. 56, pp. 1150–8.

RECKHOW, K.H. 2002. *Applications of water models: prediction uncertainty and decision-making*. Presentation at Model Uncertainty Workshop, January 2002. West Palm Beach, Fla., South Florida Water Management District.

SALTELLI, A.; CHAN, K. and SCOTT, E.M. (eds.). 2000. *Sensitivity analysis*. Chichester, UK, Wiley.

SCHWEPPE, F.C. 1973. *Uncertain dynamic systems*. Englewood Cliffs, N.J., Prentice-Hall.

SHAPIRO, H.T. 1990. The willingness to risk failure. *Science*, Vol. 250, No. 4981, p. 609.

SIMON, I.I.A. 1988. Prediction and prescription in system modelling. 15th Anniversary of IIASA, International Institute for Applied Systems Analysis. Austria, Laxenburg.

SKLAR, F.H. and HUNSAKER, C.T. 2001. The Use and uncertainties of spatial data for landscape models: an overview with examples from the Florida Everglades. In: C.T. Hunsaker, M.F. Goodchild, M.A. Friedl and T.J. Chase (eds.), *Spatial uncertainty in ecology, implications for remote sensing and GIS applications,* Chapter 2. New York, Springer.

SOROOSHIAN, S.; DUAN, Q. and GUPTA, V.K. 1993. Calibration of rainfall–runoff models: application of global optimization to the Sacramento soil moisture accounting model. *Water Resources Research*, Vol. 29, No. 4, pp. 1185–94.

SOROOSHIAN, S.; GUPTA, V.K. and FULTON, J.L. 1983. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall–runoff models: influence of calibration data variability and length on model credibility. *Water Resources Research*, Vol. 19, No. 1, pp. 251–9.

SOUTTER, M. and MUSY, A. 1999. Global sensitivity analyses of three pesticide leaching models using a Monte-Carlo approach. *Journal of Environmental Quality*, No. 28, pp. 1290–7.

SPEAR, R.C. and HORNBERGER, G.M. 1980. Eutrophication in Peel Inlet – II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research*, No. 14, pp. 43–9.

STOKEY, E. and ZECKHAUSER, R. 1977. *A primer for policy analysis*. New York, W.W. Norton.

SUTER, G.W. II. 1993. *Ecological risk assessment*. Boca Raton, Fla., Lewis.

TATTARI, S. and BARLUND, I. 2001. The concept of sensitivity in sediment yield modelling. *Physics and Chemistry of the Earth*, Vol. 26, No. 1, pp. 27–31.

VAN GRIENSVEN, A.; FRANCOS, A. and BAUWENS, W. 2002. Sensitivity analysis and auto-calibration of an integral dynamic model for river water quality. *Water Science and Technology*, Vol. 45, No. 9, pp. 325–332.

VAN HARN ADAMS, B.A. 1998. *Parameter distributions for uncertainty propagation in water quality modelling*. Doctoral Dissertation, Duke University. Durham, N.C.

VAN STRATEN, G. 1983. Maximum likelihood estimation of parameters and uncertainty in phytoplankton models. In: M.B. Beck and G. van Straten (eds.), *Uncertainty and forecasting of water quality*, New York, Springer-Verlag, pp. 157–71.

VON WINTERFELDT, D. and EDWARDS, W. 1986. *Decision analysis and behavioural research*. Cambridge, UK, Cambridge University Press.

WARWICK, J.J. and CALE, W.G. 1987. Determining likelihood of obtaining a reliable model. *Journal of Environmental Engineering*, ASCE, Vol. 113, No. 5, pp. 1102–19.

## 10. Performance Criteria


1. Introduction 293

2. Informed Decision-Making 294

3. Performance Criteria and General Alternatives 295
   - 3.1. Constraints On Decisions 296
   - 3.2. Tradeoffs 296

4. Quantifying Performance Criteria 297
   - 4.1. Economic Criteria 298
     - 4.1.1. Benefit and Cost Estimation 299
     - 4.1.2. A Note Concerning Costs 302
     - 4.1.3. Long and Short-Run Benefit Functions 303
   - 4.2. Environmental Criteria 305
   - 4.3. Ecological Criteria 306
   - 4.4. Social Criteria 308

5. Multi-Criteria Analyses 309
   - 5.1. Dominance 310
   - 5.2. The Weighting Method 311
   - 5.3. The Constraint Method 312
   - 5.4. Satisficing 313
   - 5.5. Lexicography 313
   - 5.6. Indifference Analysis 313
   - 5.7. Goal Attainment 314
   - 5.8. Goal-Programming 315
   - 5.9. Interactive Methods 315
   - 5.10. Plan Simulation and Evaluation 316

6. Statistical Summaries of Performance Criteria 320
   - 6.1. Reliability 321
   - 6.2. Resilience 321
   - 6.3. Vulnerability 321

7. Conclusions 321

8. References 322

# 10 Performance Criteria

Water resources systems typically provide a variety of economic, environmental and ecological services. They also serve a variety of purposes such as water supply, flood protection, hydropower production, navigation, recreation, waste reduction and transport. Performance criteria provide measures of just how well a plan or management policy performs. There are a variety of criteria one can use to judge and compare system performance. Some of these performance criteria may conflict with one another. In these cases tradeoffs exist among conflicting criteria and these tradeoffs must be considered when searching for the best compromise decisions. This chapter presents ways of identifying and working with these tradeoffs in the political process of selecting the best decision.

## 1. Introduction

Developing and managing water resources systems involves making decisions. Decisions are made at various levels of planning and management, even by those who are simply recommending courses of action to higher levels of an organization. The ability to make informed decisions based on scientific as well as social or political criteria is fundamental to the success of any water resources planning and management organization.

Modelling and data management tools can contribute to the information needed to make informed decisions. These tools are quantitative, but they are based on qualitative judgements about what information is and is not important to include or consider. Even the most 'objective' or 'rational' of decisions rests on subjective choices that are influenced by what decision-makers think is important, the objectives or criteria that are to be achieved and for whom, and how much they will care. One of the benefits of modelling is that differences in assumptions, in judgements over what to consider and what not to consider, or about just how accurate certain information must be, can be evaluated with respect to their influence on the decisions to be made.

Decision-makers and those who influence them are people, and people's opinions and experiences and goals differ. These differences force one to think in terms of tradeoffs. Decisions in water resources management inevitably involve making tradeoffs – compromising – among competing opportunities, goals or objectives. One of the tasks of water resources system planners or managers involved in evaluating alternative designs and management plans or policies is to identify the tradeoffs, if any, among competing opportunities, goals or objectives. It is then up to a largely political process involving all interested stakeholders to find the best compromise decision.

Measures indicating just how well different management plans and policies serve the interests of all stakeholders are typically called system performance criteria. There are many of these, some that can be quantified and some that cannot. Lack of quantification does not make those criteria less important than those that can be quantified. Those making decisions consider all criteria, qualitative as well as quantitative, as discussed in Chapter 2 among others.

If every system performance measure or objective could be expressed in the same units, and if there were only one decision-maker, then decision-making would be relatively straight forward. Such is not the case when dealing with the public's water resources.

The cost–benefit framework, used for many decades in water resources planning and management, converted the different types of impacts into a single monetary metric. Once that was done, the task was to find the plan or policy that maximized the difference between the benefits and costs. If the maximum difference between benefits and costs was positive, then that was the best plan or policy. But not all system performance criteria can be easily expressed in monetary units. Even if monetary units could be used for each objective, that in itself does not address the distributional issues of who benefits and who pays, and by how much. While all stakeholders may agree that maximizing total net benefits is a desirable objective, not everyone, if indeed anyone, will be likely to agree on how best to distribute those net benefits.

Clearly, water resources planning and management takes place in a multi-criteria environment. A key element of most problems facing designers and managers is the need to deal explicitly with multiple ecological, economic and social impacts expressed in multiple metrics that may result from management actions. Approaches that fail to recognize and explicitly include ways of handling conflict among multiple system performance measures and objectives and among multiple stakeholders are not likely to be very useful.

Successful decision-making involves creating a consensus among multiple participants in the planning and management process. These include stakeholders – individuals or interest groups who have an interest in the outcome of any management plan or policy. The relatively recent acknowledgement that stakeholders need to be fully included in the decision-making processes complicates the life of professional planners and managers. However, important sources of information come from discussion groups, public hearings, negotiations and dispute-resolution processes. It is during the debates in these meetings that information derived from the types of modelling methods discussed in this book should be presented and considered.

Eventually someone or some organization must make a decision. Even if water resources professionals view their job as one of providing technically competent options or tradeoffs for someone else to decide upon, they are still making decisions that define or limit the range of those options or tradeoffs. The importance of making informed, effective decisions applies to everyone.

## 2. Informed Decision-Making

Informed decision-making involves both qualitative thinking and analysis and quantitative modelling (Hammond et al., 1999). Qualitative thinking and analysis follows quantitative modelling. Qualitative analysis prior to quantitative modelling is useful for identifying:

- the real objectives of concern to all stakeholders (which may be other than those expressed by them)
- the likelihoods of events for which decisions are needed
- the alternatives that address and meet each objective
- the key tradeoffs among all interests and objectives involved.

When the qualitative aspects of determining the objectives and general alternative ways of meeting these objectives are weak, no amount of quantitative modelling and analysis will make up for this weakness. The identification of objectives and general alternatives are the foundation upon which water resources managers can develop appropriate quantitative models. These models will provide additional insight and definition of alternatives and their expected impacts. Models cannot identify new ideas or so called 'out-of-the-box' alternatives. Neither can they identify the best criteria that should be considered in specific cases. Only our minds can do this – individually, and then collectively.

For example, periodic municipal water supply shortages in a river basin might be reduced by building additional surface water reservoir storage capacity or by increased groundwater pumping. Assuming the objective in this case is to increase the reliability of some specified level of water supply, models can be developed to identify the tradeoffs between reservoir and pumping capacity (and cost) and increased reliability in meeting the supply target. However, these models will not consider and evaluate completely different alternatives, such as importing water by trucks, implementing water use restrictions, or water reuse, in times of drought, unless of course those options are included in the models. Someone has to think of these general alternatives before models can help identify the details: just how many trucks of what capacity, or how best to implement water reuse and for what uses.

Time needs to be taken to identify the relevant objectives and general alternatives. Clearly some preliminary screening of general alternatives should be carried out at the qualitative level, but the alternatives that remain should then be further analysed by means of quantitative modelling methods.

Creativity in identifying possible objectives and general alternatives is helped by addressing the following questions:

- What is an ideal decision?
- What does each stakeholder think other stakeholders' ideal decisions would be?
- What is to be avoided?
- What makes a great alternative, even an infeasible one, so great?
- What makes a terrible alternative so terrible?
- How would each individual's best alternative be justified to someone else?

When each manager and stakeholder has gone through such thinking, the combined set of responses may be more comprehensive and less limited by what others say or believe. They can become a basis for group discussions and consensus building.

General statements defining objectives can be converted to ones that are short and include the words maximize or minimize, e.g. minimize cost, maximize net benefits, maximize reliability, maximize water quality, maximize ecosystem biodiversity, minimize construction time, or minimize the maximum deviations from some target storage volume or target water allocation. Economic, environmental, ecological and purely physical objectives such as these can become the objective functions that drive the solutions of optimization models, as illustrated in many of the previous chapters. Social objectives should also be considered. Examples might include: maximize employment, maximize interagency coordination, maximize stakeholder participation, and minimize legal liability and the potential of future legal action and costs.

Quantitative modelling is employed to identify more precisely the design and operation of structural and non-structural alternatives that best satisfy management criteria, and to identify the impacts and tradeoffs among various performance criteria. Once such analyses have been performed, it is always wise to question whether or not the results are reasonable. Are they as expected? If not, why? If the results are surprising, are the analyses providing new knowledge and understanding, or have errors been made? How sensitive are the results to various assumptions with respect to the input data and models themselves (Chapter 9)?

Thus, the modelling process ends with some more qualitative study. Models do not replace human judgement. Humans, not quantitative models, are responsible for water resources planning and management decisions as well as the decision-making process that identifies performance criteria and general alternatives.

## 3. Performance Criteria and General Alternatives

There is a way to identify performance criteria that matter most to stakeholders (Gregory and Keeney, 1994). One can begin with very broad fundamental goals, such as public health, national as well as individual security, economic development, happiness and general well-being. Almost anyone would include these as worthy objectives. Just how these goals are to be met can be expressed by a host of other more specific objectives or criteria.

Asking 'how' any specified broad fundamental criterion or objective can be achieved usually leads to more specific system performance criteria or objectives and to the means of improving these criteria: in other words, to the general alternatives themselves. As one gets further from the fundamental objective that most if not all will agree to, there is a greater chance of stakeholder disagreement. For example, we might answer the question 'How can we achieve maximum public health?' by suggesting the maximization of surface and groundwater water quality. How? By minimizing wastewater discharges into surface water bodies and groundwater aquifers. How? By minimizing wastewater production, or by maximizing removal rates at wastewater treatment facilities, or by minimizing the concentrations of pollutants in runoff or by increasing flow augmentation or by a combination of flow augmentation and wastewater treatment. How? By increasing reservoir storage capacity and subsequent releases upstream and by upgrading wastewater treatment to a tertiary level. And so on.

Each 'how' question can have multiple answers. This can lead to a tree of branches, each branch representing a different and more specific performance criterion or an alternative way to accomplish a higher-level objective. In the example just illustrated, the answer to how to improve water quality might include a combination of water and wastewater treatment, flow augmentation, improved wastewater collection systems and reduced applications of fertilizers and pesticides on land to reduce non-point source pollutant discharges. There are others. If the lower-level objective of minimizing point source discharges had been the first objective considered, many of those other alternatives would not be identified. The more fundamental the objective, the greater will be the range of alternatives that might be considered.

If any of the alternatives identified for meeting some objective are not considered desirable, this is a good sign that there are other objectives that should be considered. For example, if flow augmentation is not desired, it could mean that in addition to water quality considerations, the regime of water flows or the existing uses of the water are also being considered, and flow augmentation may detract from those other objectives. If a stakeholder has trouble explaining why some alternative will not work, then it is a possible sign that there are other objectives and alternatives waiting to be identified and evaluated. What are they? They need to be identified. Consider them along with the others that have been already defined.

More fundamental objectives can be identified by asking the question 'why?' The answers will lead one back to increasingly more fundamental objectives. A fundamental objective is reached when the only answer to why is: 'it is what everyone really wants' or something similar. Thinking hard about 'why' will help clarify what is considered most important.

When performing qualitative exercises to identify objectives and general alternatives, enlarge the envelope and think creatively. If water needs to flow uphill, assume it can happen. If it is really worth it, engineers can do it. Lawyers can change laws. Society can and will want to adapt if it is really important. Consider for example the water that is pumped uphill at all pumped storage hydroelectric facilities, or the changing objectives over time related to managing water and ecosystems in the Mekong, Rhine and Senegal Rivers, or in the North American Great Lakes and Florida Everglades regions.

There are other traps to avoid in the planning and management process. Do not become anchored to any initial feasible or best-case or worst-case scenario. Be creative also when identifying scenarios. Consider the whole spectrum of possible events. Do not focus solely on extreme events to the exclusion of the more likely events. While toxic spills and floods bring headlines and suffering, they are not the usual more routine events that one must plan for.

It is also tempting to consider past or sunk costs when determining where additional investments should be made. If past investments were a mistake, it is only our egos that motivate us to justify those past investments by spending more money on them instead of taking more effective action.

Finally, objective targets should not be set too low. The chances of finding good unconventional alternatives is increased if targets are set high, even beyond reach. High aspirations often force individuals to think in entirely new ways. Politically it may end up that only marginal changes to the status quo will be acceptable, but it is not the time to worry about that when identifying objectives and general alternatives.

### 3.1. Constraints On Decisions

Constraints limit alternatives. Constraints are often real. There are some laws of physics and society that will limit what can be done to satisfy any performance criterion in many cases, but it is not the time to worry about them during the process of identifying objectives and general alternatives. Be creative, and do not linger on the status quo, carrying on as usual or depending on a default (and often politically risk-free) alternative. It may turn out these default or risk-free alternatives are the best available, but that can be determined later.

### 3.2. Tradeoffs

Tradeoffs are inevitable when there are multiple objectives, and multiple objectives are inevitable when there are many stakeholders or participants in the planning and management process. We all want clean water in our aquifers, streams, rivers and lakes but, if it costs money to achieve that desired quality, there are other activities or projects involving education, health care, security, or even other environmental restoration or pollution prevention and reduction efforts, which compete for those same, often limited, amounts of money.

Tradeoffs arise because we all want more of good things, and many of these good things are incompatible. We cannot have cleaner water and at the same time reduce our costs of wastewater treatment. Identifying these tradeoffs is one of the tasks of qualitative as well as quantitative analyses. Qualitatively we can identify what the tradeoffs will involve. For example, cleaner streams normally require increased costs. Just how many dinars, dollars, euros, pounds, roubles or yen it will cost to increase the minimum dissolved oxygen concentration in a specific lake by 3, or 4, or 5 mg/l is the task of quantitative analyses.

Finding a balance among all conflicting performance criteria characterizes water management decision-making. Understanding the technical information that identifies the efficient or non-dominated tradeoffs is obviously essential. Yet if the technical information fails to address the real objectives or system performance measures of interest to stakeholders, significant time and resources can be wasted in the discussions that take place in stakeholder meetings, as well as in the analyses that are performed in technical studies.

Finding the best compromise among competing decisions is a political or social process. The process is helped by having available the tradeoffs among competing objectives. Models can identify these tradeoffs among quantitative objectives but cannot go the next step, i.e. identify the best compromise decision. Models can help identify and evaluate alternatives, but they cannot take the place of human judgements that are needed to make the final decision.

Models can help identify non-dominated tradeoffs among competing objectives or system performance indicators. These are sometimes called *efficient* tradeoffs. Efficient decisions are those that cannot be altered without worsening one or more objective values. If one of multiple conflicting objective values is to be improved, some worsening of one or more other objective values will likely be necessary. Dominated or so-called inferior decisions are those in which it is possible to improve at least one objective value without worsening any of the others.

Many will argue that these dominated decisions can be eliminated from further consideration: why would any rational individual prefer such decisions when there are better ones available? However, quantitatively non-dominated or efficient solutions can be considered inferior and dominated with respect to one or more qualitative objectives, or even other quantitative objectives that were omitted from the analysis. Eliminating inferior or dominated alternatives, as defined by a set of performance criteria, from the political decision-making process may not always be very helpful. One of them could be viewed as the best by people who are considering other criteria that are either not included or, in their opinions, not properly quantified in such analyses.

Tradeoffs must not only be identified and made among conflicting outcome objectives, system performance indicators, or impacts that help define *what* should be done. Tradeoffs are also necessary among alternative processes of decision-making. Some of the most important objectives – and toughest tradeoffs – involve process decisions that establish *how* a decision is going to be made. What is, for example, the best use of time and financial resources in performing a quantitative modelling study, who should be involved in such a study, who should be advising such a study, and how should stakeholders be involved? Such questions often lie at the heart of water and other resource–use disputes and can significantly influence the trust and cooperation among all who participate in the process. They can also influence the willingness of stakeholders to support any final decision or selected management policy or plan.

## 4. Quantifying Performance Criteria

So far this chapter has focused on the critical qualitative aspects of identifying objectives and general alternatives. The remainder of the chapter will focus on quantifying various criteria and ways to use them to compare various alternatives and identify the tradeoffs among them. What is important here, however, is to realize that considerable effort is worth spending on getting the general objectives and alternatives right before spending any time on their quantification. There are no quantitative aids for this, just hard thinking, perhaps keeping in mind some of the advice just presented.

Quantification of an objective is the adoption of some quantitative (numerical) scale that provides an indicator for how well the objective would be achieved. For example, one of the objectives of a watershed conservation program might be protection or preservation of wildlife. In order to rank how various plans meet this objective, a numerical criterion is needed, such as acres of preserved habitat or populations of key wildlife species.

Quantification does not require that all objectives be described in comparable units. The same project could have a flood control objective quantified as the reduced probability of flooding or the height of the protected flood stage. It could have a regional development objective quantified as increased local income. Quantification does not require that monetary costs and benefits be assigned to all objectives.

In the following sub-sections various economic, environmental, ecological and social criteria are discussed.

## 4.1. Economic Criteria

Water resources system development and management is often motivated by economic criteria. Most of the development of water resources systems involving reservoirs, canals, hydropower facilities, groundwater-pumping systems, locks and levees has been for economic reasons. The benefits and costs of these systems can be expressed in monetary units. The criteria have been either to maximize the present value of the net benefits (total benefits less the costs) or minimize the costs of providing some purpose or service. To achieve the former involves benefit–cost analyses and to achieve the latter involves cost-effectiveness analyses.

The rationale for benefit–cost analysis is based on two economic concepts: scarcity and substitution. The former implies that the supplies of natural, synthetic and human resources are limited. Hence people are willing to pay for them. They should therefore be used in a way that generates the greatest return, i.e. they should be used efficiently. The concept of substitution, on the other hand, indicates that individuals, social groups and institutions are generally willing to trade a certain amount of one objective value for more of another.

Applied to water resources, maximization of net-benefits requires the efficient and reliable allocation, over both time and space, of water (in its two dimensions: quantity and quality) to its many uses, including hydropower, recreation, water supply, flood control, navigation, irrigation, cooling, waste disposal and assimilation, and habitat enhancement. The following example illustrates the maximization of net benefits from a multiple-purpose reservoir.

Consider a reservoir that can be used for irrigation and recreation. These two uses are not easily compatible. Recreation benefits are greater when reservoir elevations remain high throughout the summer recreation season, just when the irrigation demand would normally cause a drop in the reservoir storage level. Thus, the project has two conflicting purposes: provision of irrigation water and improvement of recreation opportunities.

Some basic principles of benefit–cost analysis can be illustrated with this example. Let $X$ be the quantity of irrigation water to be delivered to farmers each year and $Y$ the number of visitor-days of recreation use on the reservoir. Possible levels of irrigation and recreation are shown in Figure 10.1. The solid red line in Figure 10.1, termed the *production-possibility frontier* or *efficiency frontier*, is the boundary of the feasible combinations of $X$ and $Y$.

Any combination of $X$ and $Y$ within the shaded blue area can be obtained by operation of the reservoir (that is, regulating the amount of water released for irrigation and other uses). Obviously, the more of both $X$ and $Y$ the better. Thus attention generally focuses on the production-possibility frontier which comprises those combinations of $X$ and $Y$ that are *technologically efficient* in the sense that more of either $X$ or $Y$ cannot be obtained without a

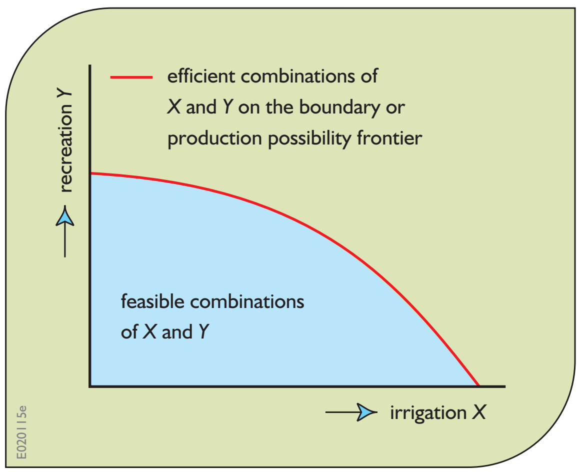



**Figure 10.1.** A two-purpose planning problem showing the feasible and efficient combinations of irrigation water $X$ and recreation visitor days $Y$.

decrease in the other. The shape and location of the production-possibility frontier is determined by the quantity of available resources (water, reservoir storage capacity, recreation facilities, etc.) and their ability to generate both $X$ and $Y$.

To continue the example, assume that a private entrepreneur owns this two-purpose reservoir in a competitive environment, i.e. there are a number of competing irrigation water suppliers and reservoir recreation sites. Let the unit market prices for irrigation water and recreation opportunities be $p_X$ and $p_Y$, respectively. Also, assume that the entrepreneur's costs are fixed and independent of $X$ and $Y$. In this case the total income is

$$I = p_X X + p_Y Y \tag{10.1}$$

Values of $X$ and $Y$ that result in fixed income levels $I_1 < I_2 < I_3$ are plotted in Figure 10.2. The value of $X$ and $Y$ that maximizes the entrepreneur's income is indicated by the points on the production-possibility frontier yielding an income of $I_3$. Incomes greater than $I_3$ are not possible.

Now assume the reservoir is owned and operated by a public agency and that competitive conditions prevail. The prices $p_X$ and $p_Y$ reflect the *value* of the irrigation water and recreation opportunities to the users. The aggregated value of the project is indicated by the user's *willingness to* pay for the irrigation and recreation outputs. In this case, this willingness to pay is $p_X X + p_Y Y$, which is equivalent to the entrepreneur's income. Private operation of the reservoir to maximize income or government operation to maximize user benefits should both, under competitive conditions, produce the same result.

When applied to water resources planning, benefit–cost analysis assumes a similarity between decision-making in the private and public sectors. It also assumes that the income resulting from the project is a reasonable surrogate for the project's social value.

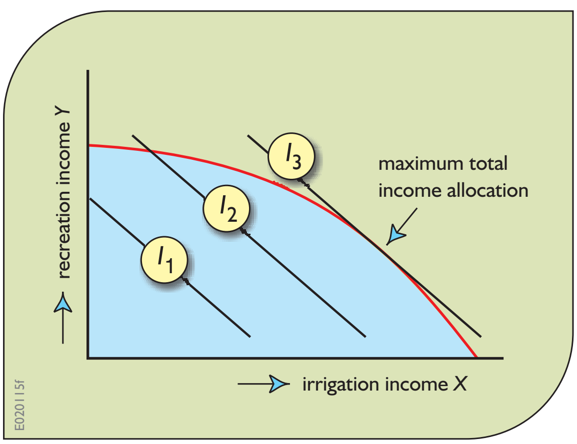



**Figure 10.2.** Two-purpose project involving tradeoffs between irrigation and recreation income and various constant total income levels, $I_1 < I_2 < I_3$.

### 4.1.1. Benefit and Cost Estimation

In benefit–cost analysis, one may need to estimate the monetary value of irrigation water, shoreline property, land inundated by a lake, lake recreation, fishing opportunities, scenic vistas, hydropower production, navigation or the loss of a wild river. The situations in which benefits and costs may need to be estimated are sometimes grouped into four categories, reflecting the way prices can be determined. These situations are:

1. Market prices exist and are an accurate reflection of marginal social values, i.e. marginal willingness to pay for all individuals. This situation often occurs in the presence of competitive market conditions. An example would be agricultural commodities that are not subsidized, that is, they do not have supported prices (some do exist!).
2. Market prices exist but for various reasons do not reflect marginal social values. Examples include price-supported agricultural crops, labour that would otherwise be unemployed, or inputs whose production generates pollution, the economic and social cost of which is not included in its price.
3. Market prices are essentially non-existent, but it is possible to infer or determine what users or consumers would pay if a market existed. An example is outdoor recreation.
4. No real or simulated market-like process can easily be conceived. This category may be relatively rare. Although scenic amenities and historic sites are often considered appropriate examples, both are sometimes privately owned and managed to generate income.

For the first three categories, benefits and costs can be measured as the *aggregate net willingness to pay* of those affected by the project. Assume, for example, that alternative water resources projects $X_1$, $X_2$, … are being considered. Let $B(X_j)$ equal the amount beneficiaries of the plan $X_j$ are willing to pay rather than forego the project. This represents the aggregate value of the project to the beneficiaries. Let $D(X_j)$ equal the amount the non-beneficiaries of plan $X_j$ are willing to pay to prevent it from being implemented. This includes the social value of the resources that will be unavailable to society if project or plan $X_j$ is implemented. The aggregate net willingness to pay $W(X_j)$ for plan $X_j$ is equal to the difference between $B(X_j)$ and $D(X_j)$,

$$W(X_j) = B(X_j) - D(X_j) \tag{10.2}$$

Plans $X_j$ can be ranked according to the aggregate net willingness to pay, $W(X_j)$. If, for example, $W(X_j) > W(X_k)$, it is inferred that plan $X_j$ is preferable or superior to plan $X_k$.

One rationale for the willingness-to-pay criterion is that if $B(X_j) > D(X_j)$, the beneficiaries could compensate the losers and everyone would benefit from the project. However, this compensation rarely happens. There is usually no mechanism established for such compensation to be paid. While many resources required for the project, such as land, labour and machinery, will be obtained and paid for, those who lose favourite scenic sites or the opportunity to use a wild river, or who must hear the noise, breathe dirtier air or suffer a loss in the value of their property because of the project are seldom compensated.

This compensation criterion also ignores the resultant income redistribution, which should be considered during the plan selection process. The compensation criterion implies that the marginal social value of income to all affected parties is the same. If a project's benefits accrue primarily to affluent individuals and the costs are borne by lower-income groups, $B(X)$ may be larger than $D(X)$ simply because the beneficiaries can pay more than the non-beneficiaries. It matters who benefits and who pays, that is, who gets to eat the pie and how much of it, as well as how big the pie is. Traditional benefit–cost analyses typically ignore this distribution issue.

In addition to these and other conceptual difficulties related to the willingness-to-pay criterion, practical measurement problems also exist. Many of the products of water resources plans are public or *collective goods*. This means that they are essentially indivisible, and once provided to any one individual it is very difficult not to provide them to others. Collective goods often also have the property that consumption by one person does not prohibit or infringe upon its consumption or use by others.

Community flood protection is an example of a public good. Once protection is provided for one individual, it is simultaneously provided for many others. As a result, it is not in an individual's self-interest to volunteer to help pay for the project by contributing an amount equal to his or her actual benefits if others are willing to pay for the project. However, if others are going to pay for the project, then individuals may exaggerate their own benefits to ensure that the project is undertaken.

Determining what benefits should be attributed to a project is not always simple and the required accounting can become rather involved. In a benefit–cost analysis, economic conditions should first be projected for a base case in which no project is implemented, and the benefits and costs are estimated for that scenario. Then the benefits and costs for each project are measured as the incremental economic impacts that occur in the economy over these baseline conditions. The appropriate method for benefit and cost estimation depends on whether or not the market prices reflecting true social values are available, or whether such prices can be constructed.

### Market Prices Equal Social Values

Consider the estimation of irrigation benefits in the irrigation–recreation example discussed in the previous section. Let $X$ be the quantity of irrigation water supplied by the project each year. If the prevailing market price $p_X$ reflects the marginal social value of water, and if that price is not affected by the project's operation, the value of the water is just $p_X X$. However, it often happens that large water projects have a major impact on the prices of the commodities they supply. In such cases, the value of water from our example irrigation district would not be based on prices before or after project implementation. Rather, the total value of the water $X$ to the users is the total amount they would be willing to pay for it.

Let $Q(p)$ be the amount of water the consumers would want to buy at a unit price $p$. For any price $p$, consumers will continue to buy the water until the value of another unit of water is less than or equal to the price $p$. The function $Q(p)$ defines what is called the demand function. As illustrated by Figure 10.3, the lower the unit price, the more water individuals are willing to buy, and the greater the demand will be.

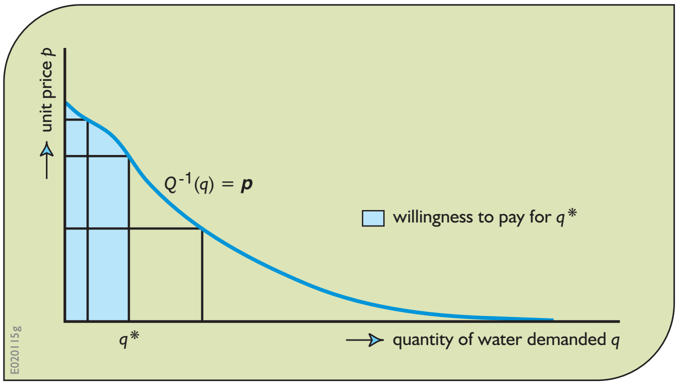



**Figure 10.3. Demand function defining how much water *Q* will be purchased for a specified unit price *p*. The shaded area represents the willingness to pay for a quantity *q*∗.**

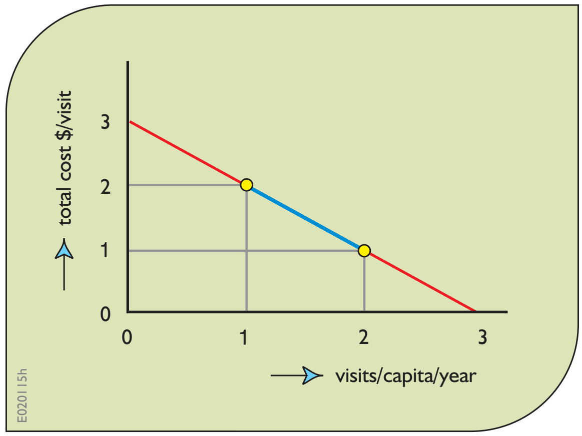



**Figure 10.4. Estimated relationship between travel cost and visits per capita.**

The willingness to pay a given unit price $p$ is defined by the area under the demand curve. As Figure 10.3 suggests, there are some who would be willing to pay a higher price for a given amount of water than others. As the unit price decreases, more individuals are willing to buy more water. The total willingness to pay for a given amount of water $q*$ is the area under the demand curve from Q = 0 to Q = $q*$:

$$\text{willingness to pay for } q* = \int_0^{q^*} Q^{-1}(q)\mathrm{d}q. \tag{10.3}$$

Consumers' willingness to pay for a product is an important concept in welfare economics.

## Market Prices Not Equal to Social Values

Market prices frequently do not truly reflect the social value of the various inputs of goods and services supplied by a water resources project. In such cases it is necessary for the planner to estimate the appropriate values of the quantities in question. There are several procedures that can be used, depending on the situation. A rather simple technique that can reach absurd conclusions if incorrectly applied is to equate the benefits of a service to the cost of supplying the service by the least expensive alternative method. Thus the benefits from hydroelectricity generation could be estimated as the cost of generating that electricity by the least-cost alternative method using solar, wind, geothermal, coal-fired, natural gas or nuclear energy sources. Clearly, this approach to benefit estimation is only valid if, were the project not adopted, the service in question would in fact be demanded at, and supplied by, the least-cost alternative method. The pitfalls associated with this method of benefit and cost estimation can be avoided if one clearly identifies reasonable with- and without project scenarios.

In other situations, simulated or imagined markets can be used to derive the demand function for a good or service and to estimate the value of the amount of that good or service generated or consumed by the project. The following hypothetical example illustrates how this technique can be used to estimate the value of outdoor recreation.

Assume a unique recreation area is to be developed, which will serve two population centres. Centre *A* has a population of 10,000, and the more distant Centre *B* has a population or 30,000. From questionnaires it is estimated that if access to the recreation area is free, then 20,000 visits per year will be made from Centre *A* at an average round-trip travel cost of $1. Similarly 30,000 visits per year will come from Centre *B*, at an average round-trip cost of $2.

The benefits derived from the proposed recreation area can be estimated from an imputed demand curve. First, as illustrated in Figure 10.4, a graph of travel cost as a function of the average number of visits per capita can be constructed. Two points are available: an average of two visits per capita (from Centre *A*) at a cost of $1/visit, and an average of one visit per capita (from Centre *B*) at a cost or $2 visit. These travel cost data are extrapolated to the ordinate and abscissa.

Even assuming that there are no plans to charge admission at the site, if users respond to an entrance fee as they respond to travel costs, then it is possible to

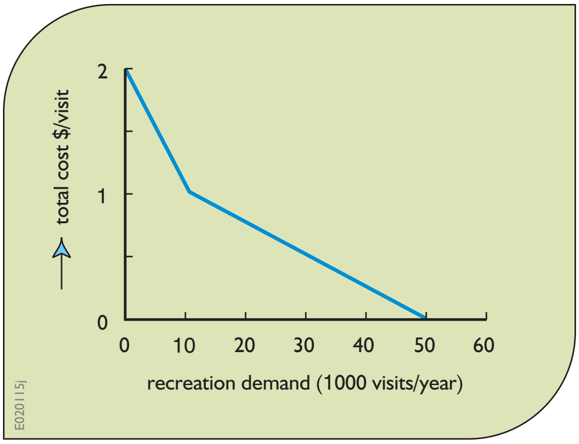



**Figure 10.5.** Demand function for recreation facility.

estimate what the user response might be if an entrance fee were to be charged. This information will provide a demand curve for recreation at the site.

Consider first a $1 admission price to be added to the travel cost for recreation. The total cost to users from population Centre *A* would then be $2 per visit. From Figure 10.4, at $2 per visit, one visit per capita is made; hence, 10,000 visits per year can be expected from Centre *A*. The resulting cost to users from Centre *B* is $3 per visit, and hence, from Figure 10.4 no visits would be expected. Therefore, one point on the demand curve (10,000 visits at $1) is obtained. Similarly, it can be inferred that at a $2 admission price there will be no visits from either centre. A final point, corresponding to a zero admission price (no added costs) is just the expected site attendance (20,000 + 30,000 = 50,000 visits). The resulting demand curve is shown in Figure 10.5. Recreation total willingness-to-pay benefits are equal to the area under the demand curve, or $35,000. Assuming no entrance fee is charged, this amounts to $0.70 per visitor-day based on the expected 50,000 visits.

Obviously, this example is only illustrative. The cost of travel time (which differs for each population centre) and the availability of alternative sites must be included in a more detailed analysis.

### No Market Processes

In the absence of any market-like process (real or simulated), it is difficult to associate specific monetary values with benefits. The benefits associated with aesthetics and with many aspects of environmental and ecosystem quality have long been considered difficult to quantify in monetary terms. Although attempts (some by highly respected economists) have been made to express environmental benefits in monetary terms, the results have had limited success. In most regions in the world, water resources management guidelines, where they exist, do not encourage the assignment of monetary values to these criteria. Rather the approach is to establish environmental and ecological requirements or regulations. These constraints are to be met, perhaps while maximizing other economic benefits. Legislative and administrative processes rarely, if ever, determine the explicit benefits derived from meeting these environmental and ecological requirements or regulations.

To a certain extent, environmental (including aesthetic) objectives, if quantified, can be incorporated into a multi-objective decision-making process. However, this falls short of the assignment of monetary benefits that is often possible for the first three categories of benefits.

### 4.1.2. A Note Concerning Costs

To be consistent in the estimation of net benefits from water resources projects, cost estimates should reflect opportunity costs, the value of resources in their most productive alternative uses. This principle is much easier stated than implemented, and as a result true opportunity costs are seldom included in a benefit–cost analysis. For example, if land must be purchased for a flood-control project, is the purchase price (which would typically be used in a benefit–cost analysis) the land's opportunity cost? Suppose that the land is currently a natural area and its alternative use is as a nature preserve and camping area. The land's purchase price may be low, but this price is unlikely to reflect the land's true value to society. Furthermore, assume that individuals who would otherwise be unemployed are hired to work on the land. The opportunity cost for such labour is the marginal value of leisure forgone, since there is no alternative productive use of those individuals. Yet the labourers must be paid and their wages must be included in the budget for the project.

The results of rigorous benefit–cost analyses seldom dictate which of competing water resources projects and plans should be implemented. This is in part because of the multi-objective nature of the decisions. One must consider environmental impacts, income redistribution effects and a host of other local, regional and national goals, many of which may be non-quantifiable.

Other important considerations are the financial, technical and political feasibilities of alternative plans. Particularly important when a plan is undertaken by government agencies is the relative political and legal clout of those who support the plan and those who oppose it. Still, a plan's economic efficiency is an important measure of its value to society and often serves as an indicator of whether it should be considered at all.

### 4.1.3. Long and Short-Run Benefit Functions

When planning the capacities and target values associated with water resources development projects, it is often convenient to think of two types of benefit functions: what economists call long-run benefits and short-run benefits. In *long-run planning*, the capacities of proposed facilities and the target allocations of flows or storage volumes to alternative uses are unknown decision-variables. The values of these variables are to be determined in a way that achieves the most beneficial use of available resources, even when the available resources vary in magnitude and, hence, do not always equal the targets over time. In *short-run planning*, the capacity of facilities and the allocation targets of water promised to alternative uses are specified, i.e. they are known. The problem is one of managing or operating a given or proposed system under varying supply conditions.

For example, if a water-using firm is interested in building a factory along a river, of interest to those designing the factory is the amount of water the factory can expect to get. This in part may dictate its capacity, the number of employees hired and so on. On the other hand, if the factory already exists, the likely issue is how to manage or operate the factory when the water supply varies from the target levels that were (and perhaps still are) expected.

Long-run benefits are those benefits obtained if all target allocations are met. Short-run benefits are the benefits one actually obtains by operating a system to best use the available resources after capacities and target values have been fixed by prior decisions. If the resources available in the short-run are those that can meet all the targets that were established when long-run decisions were made, then estimated long-run benefits can be achieved. Otherwise the benefits actually obtained may differ from those expected. The goal is to determine the values of the long-run decision-variables in a way that maximizes the present value of all the short-run benefits obtained given the varying water supply conditions.

The distinction between long-run and short-run benefits can be illustrated by considering again a potential water user at a particular site near a water body. Assume that the long-run net benefits associated with various target allocations of water to that use can be estimated. These long-run net benefits are those that will be obtained if the actual allocation $Q$ equals the target allocation $Q^T$. This long-run net benefit function can be denoted as $B(Q^T)$. Next assume that for various fixed values of the target $Q^T$ the actual net benefits derived from various allocations $Q$ can be estimated. These short-run benefit functions $b(Q|Q^T)$ of allocations $Q$ given a target allocation $Q^T$ are dependent on the target $Q^T$ and, obviously, on the actual allocation $Q$. The relationship between the long-run net benefits $B(Q^T)$ and the short-run net benefits $b(Q|Q^T)$ for a particular target $Q^T$ is illustrated in Figure 10.6.

The long-run function $B(Q^T)$ in Figure 10.6 reflects the benefits that users receive when they have adjusted their plans in anticipation of receiving an allocation equal to the target $Q^T$ and actually receive it. The short-run benefits function specifies the benefits users actually

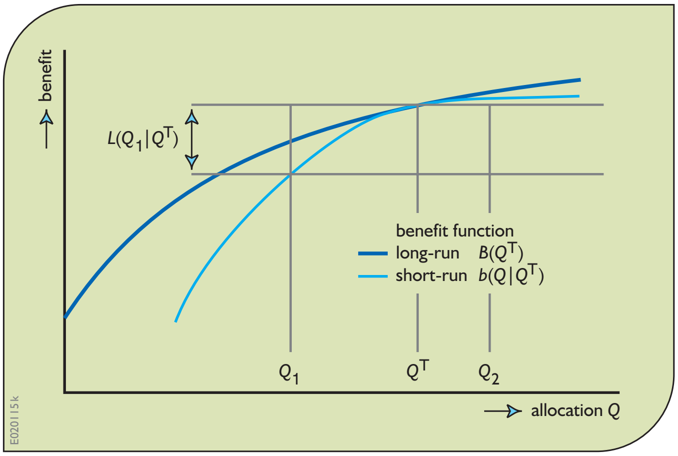



**Figure 10.6.** Long-run and short-run benefit functions, together with the loss, $L(Q_1|Q^T)$, associated with a deficit allocation $Q_1$ and target allocation $Q^T$.

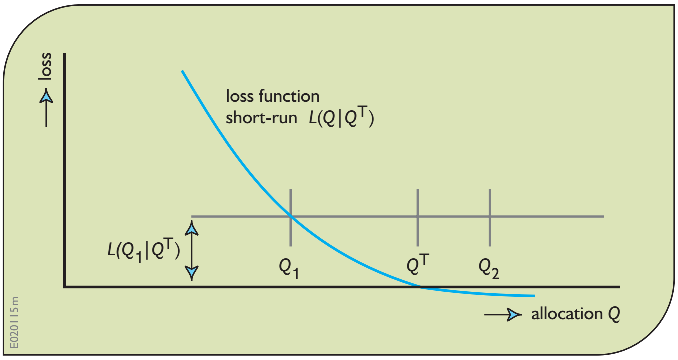



**Figure 10.7.** Short-run loss function, $L(Q|Q^T)$, associated with an allocation $Q$ and target allocation $Q^T$.

receive when a particular allocation is less (e.g. $Q_1$) or more (e.g. $Q_2$) than the anticipated allocation, $Q^T$, and they cannot completely adjust their plans to the resulting deficit or surplus.

Clearly the short-run benefits associated with any allocation cannot be greater than the long-run benefits obtainable had the firm planned or targeted for that allocation. The short-run benefit function is always going to be under, or tangent to, the long-run benefit function, as shown in Figure 10.6. In other words the short-run benefits $b(Q\,|\,Q^T)$ will never exceed the long-run benefits $B(Q)$ that could be obtained if the target $Q^T$ were equal to the allocation $Q$. When the target $Q^T$ equals the allocation $Q$, the values of both functions are equal.

Flipping the short run benefit function upside-down along a horizontal axis running through the long-run benefit function at the target $Q^T$ defines the short-run loss function, as illustrated in Figure 10.7. The short-run loss of any actual allocation $Q$ equals the long-run benefit of the target allocation $Q^T$ minus the short-run benefit of the actual allocation, $Q$:

$$L(Q\,|\,Q^T) = B(Q^T) - b(Q\,|\,Q^T) \tag{10.4}$$

This function defines the losses that occur when the target allocation cannot be met. When the actual allocation equals the target allocation, the short-run loss is zero. It is possible there might be short-run gains or benefits if there is a surplus allocation over the target allocation, possibly for a limited range of excess allocations.

The short-run benefit function, or its corresponding loss function, usually depends on the value of the target allocation. However, if the short-run losses associated with any deficit allocation $(Q^T - Q_1)$ or surplus allocation $(Q_2 - Q^T)$ are relatively constant over a reasonable range of targets, it may not be necessary to define the loss as a function of the target $Q^T$. In this case the loss can be defined as a function of the deficit $D$ and/or as a function of the surplus (excess) $E$. Both the deficit $D$ or excess $E$ can be defined by the constraint

$$Q = Q^T - D + E \tag{10.5}$$

Denote the loss function for a deficit allocation as $L^D(D)$ and for a surplus allocation, $L^E(E)$. As indicated above, the latter may be a negative loss – a gain – at least for some range of $E$, as shown in Figure 10.7.

The costs of the capacity of many components or multipurpose projects are not easily expressed as functions of the targets associated with each use. For example, the capacity, $K$, of a multipurpose reservoir is not usually equal to, or even a function of, its recreation level target or its active or flood storage capacities. The costs of its total capacity, $C(K)$, are best defined as functions of that total capacity. If expressed as an annual cost, this would include the annual amortized capital costs as well as the annual operation, maintenance and repair costs.

Assuming that the benefit and loss functions reflect annual benefits and losses, the annual net benefits, $NB$, from all projects $j$ is the sum of each project's long-run benefits $B_j(T_j)$ that are functions of their targets, $T_j$, less short-run losses $L_j(Q_j\,|\,T_j)$ and capacity costs $C_j(K_j)$:

$$NB = \sum_j [B_j(T_j) - L_j(Q_j\,|\,T_j) - C_j(K_j)] \tag{10.6}$$

The monetary net benefits accrued by each group of water users can also be determined so that the income redistribution effects of a project can be evaluated.

The formulation of benefit functions as either long or short-run is, of course, a simplification of reality. In reality, planning takes place on many time scales, and for each time scale one could construct a benefit function. Consider the planning problems of farmers. In the very long run, they decide whether or not to own farms, and if so how big they are to be in a particular area. On a shorter time scale, farmers allocate their resources to different activities depending on what products they are producing and on the processes used to produce them. Different activities, of course, require capital investments in farm machinery, storage facilities, pipes, pumps and the like, some of which cannot easily be transferred to other uses.

At least on an annual basis, most farmers re-appraise these resource allocations in light of the projected market prices of the produced commodities and the availability and cost of water, energy, labour and other required inputs. Farmers can then make marginal adjustments in the amounts of land devoted to different crops, animals and related activities within the bounds allowed by available resources, including capital and labour. At times during any growing season some changes can be made in response to changes in prices and the actual availability of water.

If the farmers frequently find that insufficient water is available in the short run to meet livestock and crop requirements, then they will reassess and perhaps change their long-run plans. They may shift to less water-intensive activities, seek additional water from other sources (such as deep wells), or sell their farms (and possibly water rights) and engage in other activities. For the purposes of modelling, however, this planning hierarchy can generally be described by two levels, denoted as long and short run. The appropriate decisions included in each category will depend on the time scale of a model.

These long-run and short-run benefit and loss functions are applicable to some water users, but not all. They may apply in situations where benefits or losses can be attributed to particular allocations in each of the time periods being modelled. They do not apply in situations where the benefits or losses result only at the end of a series of time periods, each involving an allocation decision. Consider irrigation, for example. If each growing season is divided into multiple periods, then the benefits derived from each period's water allocation cannot be defined independently of any other period's allocation, at least very easily. The benefits from irrigation come only when the crops are harvested, for example, in the last period of each growing season. In this case some mechanism is needed to determine the benefits obtained from a series of interdependent allocations of water over a growing season, as will be presented in the next chapter.

## 4.2. Environmental Criteria

Environmental criteria for water resources projects can include water quantity and quality conditions. These are usually expressed in terms of targets or constraints for flows, depths, hydroperiods (duration of flooding), storage volumes, flow or depth regimes, and water quality concentrations that are considered desirable for aesthetic or public health reasons or for various ecosystem habitats. These constraints or targets could specify desired minimum or maximum acceptable ranges, or rates of changes, of these values, either for various times within each year or over an *n*-year period.

Water quality constituent concentrations are usually expressed in terms of some maximum or minimum acceptable concentration, depending on the particular constituents themselves and the intended uses of the applicable water body. For example, phosphorus would normally have a maximum permissible concentration, while dissolved oxygen would normally have a minimum acceptable concentration. These limiting concentrations and their specified reliabilities are often based on standards established by national or international organizations. As standards, they are viewed as constraints. These standards could also be considered as targets, and the maximum or average adverse deviations from these standards or targets could be a system performance measure.

Some water quality criteria may be expressed in qualitative terms. Qualitative quality criteria can provide a 'fuzzy' limit on the concentration of some constituent in the water. Such criteria might be expressed as, for example, 'the surface water shall be virtually free from floating petroleum-derived oils and non-petroleum oils of vegetable or animal origin.' Fuzzy membership functions can define what is considered 'virtually free,' and these can be included as objectives or constraints in models (as discussed in Chapter 5).

Environmental performance criteria can vary depending on the specific sites and on the intended uses of water at that site. They should be designed to assess, or define, the risks of adverse impacts on the health of humans and aquatic life from exposure to pollutants.

Environmental performance criteria of concern to water resources planners and managers can also relate to recreational and land use activities. These typically address hydrological conditions, such as streamflows or lake or reservoir storage volume elevations during specified times, or land use activities on watersheds. For example, to increase the safety of sailors and individuals fishing downstream of hydropower reservoirs, release rules may have to be altered to reduce the rate of flow increase that occurs during peak power production. Performance criteria applicable to the adverse environmental impacts of sediment loads – such as those caused by logging or construction activities, or to the impact of nutrient loads in the runoff, perhaps from agricultural and urban areas – are other examples.

## 4.3. Ecological Criteria

Criteria that apply to aquatic ecosystems involve both water quantity and quality and are often compatible with environmental criteria. It is the time-varying regimes of water quantities and qualities, not minimum or maximum values, that benefit and have impacts upon ecosystems. It is not possible to manage water and its constituent concentrations in a way that maximizes the health or well-being, however measured, of all living matter in an ecosystem. When conditions favour one species that feeds on another, it is hard to imagine how the latter can be as 'healthy' as the former. The conditions that favour one species group may not favour another. Hence, variation in habitat conditions is important for the sustainability of resilient biodiverse ecosystems.

While ecosystem habitats exhibit more diversity when hydrological conditions vary, as in nature, than when they are relatively constant, hydrological variation is often not desired by most human users of water resources systems. Reducing hydrological variation and increasing the reliability of water resources systems has often been the motivating factor in the design and operation of hydraulic engineering works.

The states of ecologial habitats are in part functions of how water is managed. One way to develop performance indicators of ecosystems is to model the individual species that they are composed of, or at least a subset of, important indicator species. This is often very difficult. Alternatively, one can define habitat suitability indices for these important indicator species. This requires, first, selecting the indicator species representative of each particular ecosystem, second, identifying the hydrological attributes that affect the well-being of those indicator species during various stages of their life cycles, and third, quantifying the functional relationships between the well-being of those species and values of the applicable hydrological attributes, usually on a scale from 0 to 1. A habitat suitability value of 1 is considered an ideal condition. A value of 0 is considered to be very unfavourable.

Examples of hydrological attributes that affect ecosystems could include flow depths, velocities, constituent concentrations and temperatures, their durations or the rate of change in any of those values over space and/or time at particular times of the year. In wetlands, the hydrological attributes could include the duration of inundation (hydroperiod), time since last drawdown below some threshold depth, the duration of time below or above some threshold depth, and time rates of change in depth. The applicable attributes themselves, or perhaps just their functional relationships, can vary depending on the time of year and on the stage of species development.

Figure 10.8 illustrates three proposed habitat suitability indices for periphyton (algae) and fish in parts of the Everglades region of southern Florida in the United States. All are functions of hydrological attributes that can be managed. Shown in this figure is the impact of hydroperiod duration on the habitat of three different species of periphyton located in different parts of the Everglades, and the impact of the duration of the hydroperiod as well as the number of years since the last dry period on a species of fish.

There are other functions that would influence the growth of periphyton and fish, such as the concentrations of phosphorus or other nutrients in the water. These are not shown. Figure 10.8 merely illustrates the construction of habitat suitability indices.

There are situations where it may be much easier and more realistic to define a range of some hydrological attribute values as being ideal. Consider fish living in streams or rivers for example. Fish desire a variety of depths, depending on their feeding and spawning activities. Ideal trout habitat, for example, requires both deeper pools and shallower riffles in streams. There is no ideal depth or velocity, or even a range of depths or velocities above or below some threshold value. In such cases, it is possible to divide the hydrological attribute scale into discrete ranges and identify the ideal fraction of the entire stream or river reach or wetland that ideally would be within each discrete range. This exercise will result in a function resembling a probability density function. The total area under the function is 1. (The first and last segments of the function can represent anything less or greater than some discrete value, where appropriate, and

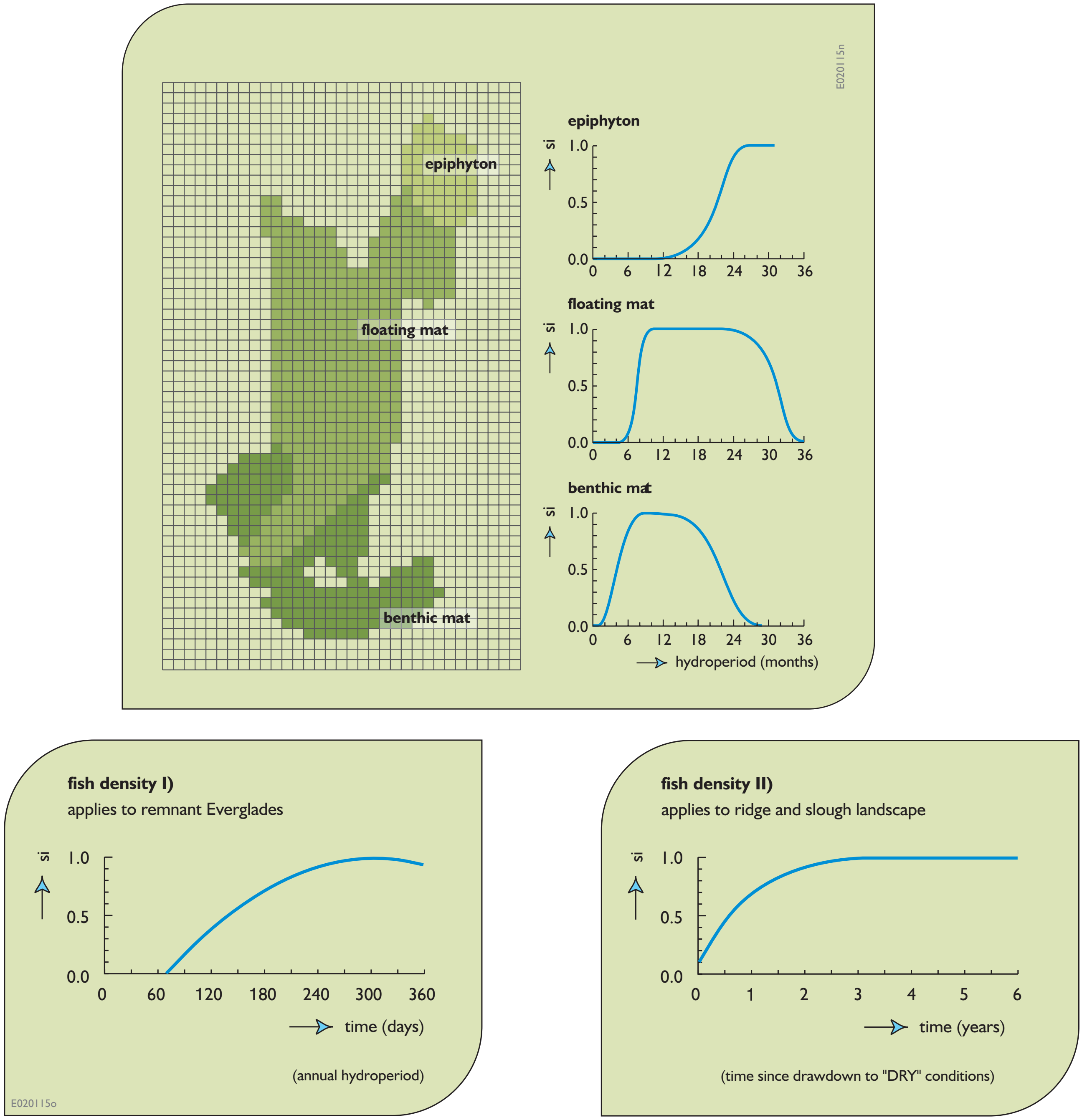



**Figure 10.8.** Some proposed habitat suitability indices (SI) for three types of periphyton (algae) and a species of fish in portions of the Everglades region in southern Florida of the United States.

if so the applicable segments are understood to cover those ranges.) Such a function is shown in Figure 10.9. Figure 10.9 happens to be a discrete distribution, but it could have been a continuous one as well.

Any predicted distribution of attribute values resulting from a simulation of a water management policy can be compared to this ideal distribution, as is shown in Figure 10.10. The fraction of overlapping areas of both distributions is an indication of the suitability of that habitat for that indicator species. In Figure 10.10, this is the red shaded area under the blue curve, divided by the total red shaded area.

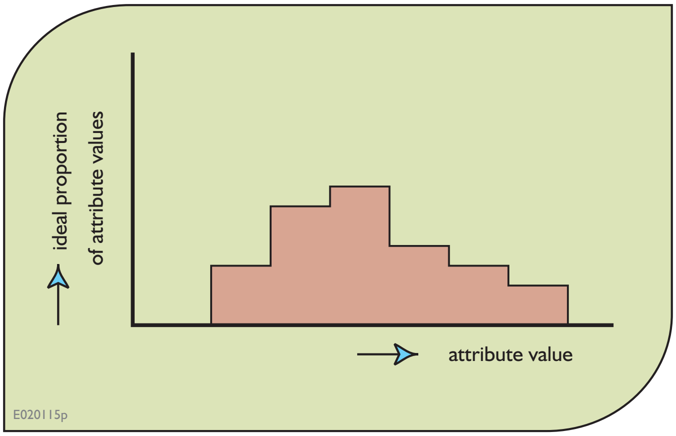



**Figure 10.9.** Ideal range of values of a hydrological attribute for a particular component of an ecosystem.

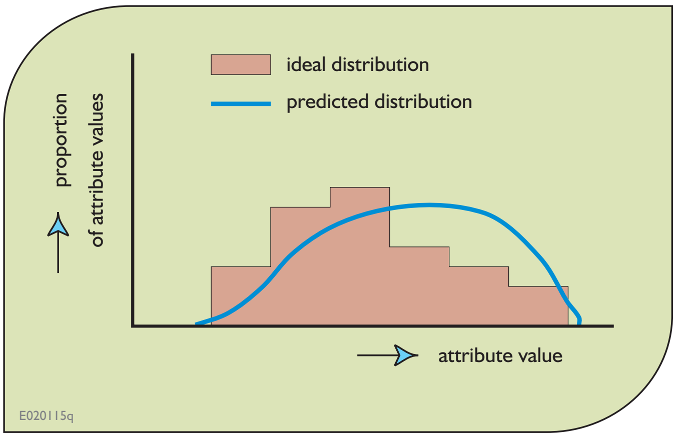



**Figure 10.10.** Predicted (simulated indicated in blue) and ideal (indicated in red) distributions of attribute values for some ecosystem indicator species. The fraction of the area common to both distributions is a measure of habitat suitability.

$$\text{Habitat suitability} = \text{Fraction of area under the ideal and simulated distributions of attribute values} \qquad (10.7)$$

To identify a representative set of indicator species of any ecosystem, the hydrological attributes or 'stressors' that affect those indicator species, and finally the specific functional relationships between the hydrological attributes and the habitat suitability performance indicators, is not a trivial task. The greater the number of experienced individuals involved in such an exercise, the more difficult it may be to reach some agreement or consensus. This just points to the complexity of ecosystems and the difficult task of trying to simplify it in order to define habitat suitability performance criteria. However, once identified, these habitat suitability performance criteria can give water resources planners and managers an admittedly incomplete but at least relative indication of the ecosystem impacts of alternative water management policies or practices.

The use of these habitat suitability functions along with other performance criteria in optimization and simulation models will be discussed later in this chapter.

## 4.4. Social Criteria

Social performance criteria are often not easily defined as direct functions of hydrological attributes. Most social objectives are only indirectly related to hydrological attributes or other measures of water resources system performance. Economic, environmental and ecological impacts resulting from water management policies affect people directly. One social performance criterion that has been considered in some water resources development projects, especially in developing regions, has been employment. Where employment is considered important, alternatives that provide more jobs may be preferred to those that use more heavy machinery in place of labour, for example.

Another social performance criterion is human settlement displacement. The number of families that must move from their homes because of, for example, floodplain restoration or reservoir construction is always of concern. These impacts can be expressed as a function of the extent of floodplain restoration or reservoir storage capacity, respectively. Often the people most affected are in the lower income groups, and this raises legitimate issues of justice and equity. Human resettlement impacts have both social and economic dimensions.

Social objectives are often the more fundamental objectives discussed earlier in this chapter. Asking 'why' identifies them. Why improve water quality? Why prevent flood damage? Why, or why not, build a reservoir? Why restore a floodplain or wetland? Conversely, if social objectives are first identified, by asking and then answering 'how' they can be achieved, this usually results in the identification of economic, environmental and ecological objectives more directly related to water management.

The extent of press coverage or of public interest and participation in the planning and evaluation processes can also be an indicator of social satisfaction with water management. It is in times of social stress due to, for example, floods, droughts, or disease caused by water-borne bacteria, viruses and pollutants that social impacts become important, that press coverage and public involvement increases. (Public interest also increases when there is a lot of money to be spent, but this is often a result of substantial public interest as well.) It is a continuing challenge to actively engage an often uninterested public in water management planning at times when there are no critical water management impacts being felt and not a lot of money is being spent. Yet this is just the time when planning for more stressful conditions should take place.

## 5. Multi-Criteria Analyses

Given multiple performance criteria measured in multiple ways, how can one determine the best decision, i.e. the best way to develop and manage water? Just what is best, or as some put it, rational? The answer to these questions will often differ depending on who is being asked. There is rarely an alternative that makes every interest group or affected stakeholder the happiest. In such cases we can identify the efficient tradeoffs among the objective values each stakeholder would desire. In this section, some ways of identifying efficient tradeoffs are reviewed. These methods of multi-criteria or multi-objective analyses are not designed to identify the best solution, but only to provide information on the tradeoffs between given sets of quantitative performance criteria. Again, any final decision will be based on qualitative as well as this quantitative information in a political process, not in a computer.

Even if the same, say monetary, units of measure can be used for each planning or management objective, it is not always possible to identify the maximum net benefit alternative. Consider for example a water resources development project to be designed to maximize net economic benefits. In the United States, this is sometimes designated as *NED*: national economic development. That is one objective. A second objective is to distribute the costs and benefits of the project in an equitable way. Both objectives are measured in monetary units. While everyone may agree that the biggest pie (i.e. the maximum net benefits) should be obtained, subject to various environmental and ecological constraints perhaps, not everyone is likely to agree as to how that pie should be divided up among all the hungry stakeholders. Again, issues of equity and justice involve judgements, and the job of water resources planners and managers and elected politicians, is to make good ones. The result is often a solution that does not maximize net economic benefits. Requiring all producers of wastewater effluent to treat their wastes to the best practical level before discharging the remainder, regardless of the assimilative capacity of the receiving bodies of water, is one example of this compromise between economic and social criteria.

The irrigation–recreation example presented earlier in this chapter illustrates some basic concepts in multi-objective planning. As indicated in Figure 10.1, one of the functions of multi-objective planning is the identification of plans that are technologically efficient. These are plans that define the production-possibility frontier. Feasible plans that are not on this frontier are inferior in the sense that it is always possible to identity alternatives that will improve one or more objective values without making others worse.

Although the identification of feasible and efficient plans is seldom a trivial matter, it is conceptually straightforward. Deciding which of these efficient plans is the best is quite another matter. Social welfare functions that could provide a basis for selection are impossible to construct, and the reduction of multiple objectives to a single criterion (as in Figure 10.2) is seldom acceptable in practice.

When the various objectives of a water resources planning project cannot be combined into a single scalar objective function, a single objective management model may no longer be appropriate. Rather, a *vector optimization* formulation of the problem may be applicable. Let the vector $\boldsymbol{X}$ represent the set of unknown decision-variable values that are to be determined, and $Z_j(\boldsymbol{X})$ be a performance criterion or objective that is to be maximized. Each performance criterion or objective $j$ is a function of these unknown decision-variable values. Assuming that all $J$ objectives $Z_j(\boldsymbol{X})$ are to be maximized, the model can be written maximize $[Z_1(\mathbf{X}), Z_2(\mathbf{X}), \ldots, Z_j(\mathbf{X}), \ldots, Z_J(\mathbf{X})]$

subject to:

$$g_i(\mathbf{X}) = b_i \quad i = 1, 2, \ldots, m \tag{10.8}$$

The objective in Equation 10.8 is a vector consisting of $J$ separate objectives. The $m$ constraints $g_i(\mathbf{X}) = b_i$ define the feasible region of solutions.

The vector optimization model is a concise way of formulating a multi-objective problem, but it is not very useful in solving it. In reality, a vector can be maximized only if it can be reduced to a scalar. Thus, the multi objective planning problem defined by Model 10.8 cannot, in general, be solved without additional information. Various ways of solving this multi-objective model are discussed in the following sub-sections.

The goal of multi-objective modelling is the generation of a set of technologically feasible and efficient plans. An efficient plan is one that is not dominated.

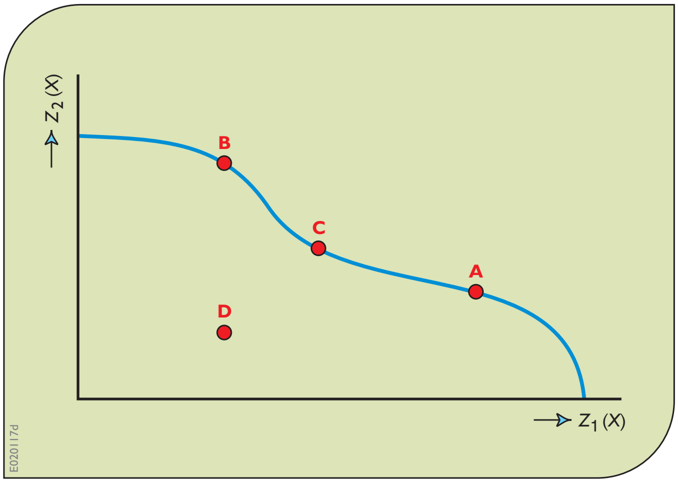



**Figure 10.11.** Four discrete plans along with a continuous efficiency frontier associated with two objectives, $Z_1$ and $Z_2$. A pair-wise comparison of plans or objectives may not identify all the non-dominated plans. All objectives should be considered before declaring a plan inferior.

## 5.1. Dominance

A plan $\mathbf{X}$ dominates all others if it results in an equal or superior value for all objectives, and at least one objective value is strictly superior to those of each other plan. In symbols, assuming that all objectives $j$ are to be maximized, plan alternative $i$, $\mathbf{X}_i$, dominates if

$$Z_j(\mathbf{X}_i) \geq Z_j(\mathbf{X}_k) \quad \text{for all objectives } j \text{ and plans } k \tag{10.9}$$

and for each plan $k \neq i$ there is at least one objective $j^*$ such that

$$Z_{j^*}(\mathbf{X}_i) > Z_{j^*}(\mathbf{X}_k) \tag{10.10}$$

It is seldom that one plan dominates all others. If it does, choose it! It is more often the case that different plans will dominate all plans for different objectives. However, if there exists two plans $k$ and $h$ such that $Z_j(\mathbf{X}_k) \geq Z_j(\mathbf{X}_h)$ for all objectives $j$ and for some objective $j^*$, $Z_{j^*}(\mathbf{X}_k) > Z_{j^*}(\mathbf{X}_h)$, then plan $k$ dominates plan $h$ and plan $\mathbf{X}_h$ can be dropped from further consideration, assuming of course that all objectives are being considered. If some objectives are not, or cannot be, included in the analysis, such inferior plans with respect to the objectives that are included in the analysis should not be rejected from eventual consideration. Dominance analysis can only deal with the objectives being explicitly considered.

Dominance analysis requires that participants in the planning and management process specify the objectives that are to be maximized or minimized. It does not require the assessment of the relative importance of each objective. Non-inferior, efficient, or non-dominated solutions are often called Pareto optimal because they satisfy the conditions proposed by the nineteenth-century Italian economist Vilfredo Pareto, namely that to improve the value of any single objective, one should have to accept a diminishment of at least one other objective.

Consider for example three alternatives $A$, $B$ and $C$. Assume, as shown in Figure 10.11, that plan $C$ is inferior to plan $A$ with respect to objective $Z_1(\mathbf{X})$ and also inferior to plan $B$ with respect to objective $Z_2(\mathbf{X})$. Plan $C$ might still be considered the best with respect to both objectives $Z_1(\mathbf{X})$ and $Z_2(\mathbf{X})$. While plan $C$ could have been inferior to both $A$ and $B$, as is plan $D$ in Figure 10.11, it should not necessarily be eliminated from consideration just on the basis of a pair-wise comparison. In fact, plan $D$, even though inferior with respect to both objectives $Z_1(\mathbf{X})$ and $Z_2(\mathbf{X})$, might be the preferred plan if another objective were included.

Two common approaches for identifying non-dominated plans that together identify the efficient trade-offs among all the objectives $Z_j(\mathbf{X})$ in the Model 10.8 are the *weighting* and *constraint* methods. Both methods require numerous solutions of a single-objective management model to generate points on the objective functions' production-possibility frontier.

### 5.2. The Weighting Method

The weighting approach involves assigning a relative weight to each objective to convert the objective vector (in Equation 10.8) to a scalar. This scalar is the weighted sum of the separate objective functions. The multi-objective Model 10.8 becomes

$$\text{maximize } Z = [w_1 Z_1(\boldsymbol{X}) + w_2 Z_2(\boldsymbol{X}) \cdots + w_j Z_j(\boldsymbol{X}) \cdots + w_J Z_J(\boldsymbol{X})]$$

subject to:

$$g_i(\boldsymbol{X}) = b_i \quad i = 1, 2, \ldots, m \tag{10.11}$$

where the non-negative weights $w_j$ are specified constants. The values of these weights $w_j$ are varied systematically, and the model is solved for each combination of weight values to generate a set of technically efficient (or non-inferior) plans.

The foremost attribute of the weighting approach is that the tradeoffs or marginal rate of substitution of one objective for another at each identified point on the objective function's production-possibility frontier is explicitly specified by the relative weights. The marginal rate of substitution between any two objectives $Z_j$ and $Z_k$, at a specified constant value of $\boldsymbol{X}$, is

$$-[\mathrm{d}Z_j / \mathrm{d}Z_k] \,|_{x\,=\,\text{constant}} = w_k / w_j \tag{10.12}$$

This applies when each of the objectives is continuously differentiable at the point $\boldsymbol{X}$ in question. This is illustrated for a two-objective maximization problem in Figure 10.12.

These relative weights can be varied over reasonable ranges to generate a wide range of plans that reflect different priorities. Alternatively, specific values of the weights can be selected to reflect preconceived ideas of the relative importance of each objective. It is clear that the prior selection of weights requires value judgements. If each objective value is divided by its maximum possible value, then the weights can range from 0 to 1 and sum to 1, to reflect the relative importance given to each objective.

For many projects within developing countries, these weights are often estimated by the agencies financing the development projects. The weights specified by these agencies can, and often do, differ from those implied by

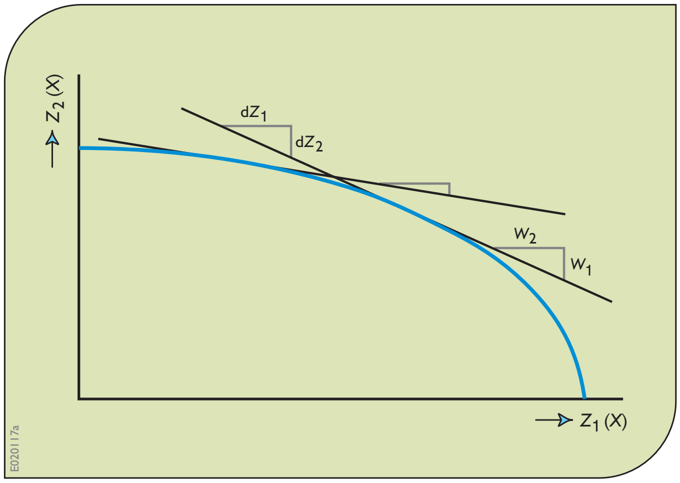



**Figure 10.12.** The efficiency frontier between two objectives, $Z_1(\boldsymbol{X})$ and $Z_2(\boldsymbol{X})$, showing the reduction in one objective, say $Z_1(\boldsymbol{X})$, as the relative weight, $w_2$, associated with the other objective, increases.

national or regional policy. But regardless of who does it, the estimation of appropriate weights requires a study of the impacts on the economy, society and development priorities involved.

Fortunately here we are not concerned with finding the best set of weights, but merely using these weights to identify the efficient tradeoffs among the conflicting objectives. After a decision is made, the weights that produce that solution might be considered the best, at least under the circumstances and at the time when the decision is made. They will probably not be the weights that will apply in other places in other circumstances at other times.

A principal disadvantage of the weighting approach is that it cannot generate the complete set of efficient plans unless the efficiency frontier is strictly concave (decreasing slopes) for maximization, as it is in Figures 10.1 and 10.12. If the frontier, or any portion of it, is convex, then only the endpoints of the convex region will be identified using the weighting method when maximizing, as illustrated in Figure 10.13.

### 5.3. The Constraint Method

The constraint method for multi-objective planning can produce the entire set of efficient plans for any shape of efficiency frontier, including that shown in Figure 10.13,

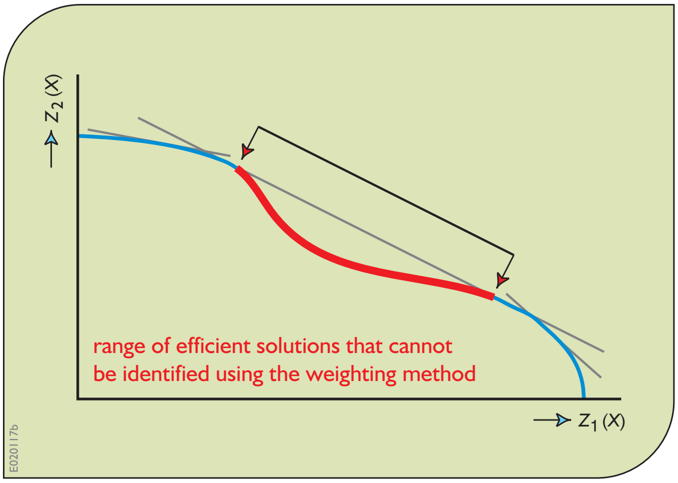



**Figure 10.13.** An efficiency frontier that cannot be completely identified in its convex region using the weighting method when objectives are being maximized. Similarly for concave regions when objectives are being minimized.

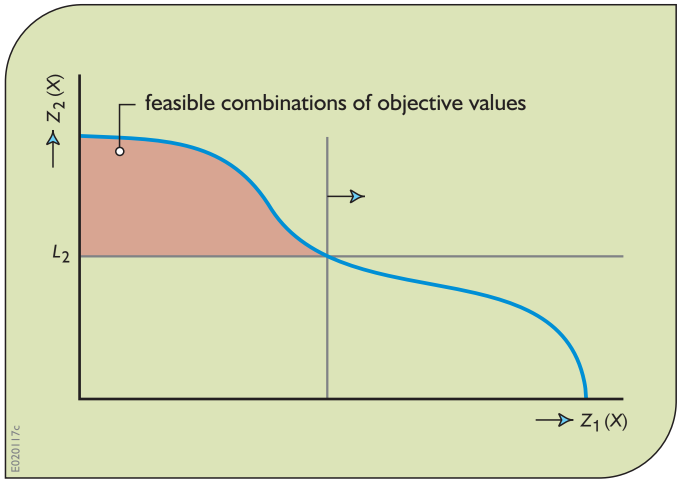



**Figure 10.14.** The constraint method for identifying the efficiency frontier by maximizing $Z_1(\boldsymbol{X})$ while constraining $Z_2(\boldsymbol{X})$ to be no less than $L_2$.

assuming there are tradeoffs among the objectives. In this method, one objective, say $Z_k(\boldsymbol{X})$, is maximized subject to lower limits, $L_j$, on the other objectives, $j \neq k$. The solution of the model, corresponding to any set of feasible lower limits $L_j$, produces an efficient alternative if the lower bounds on the other objective values are binding.

In its general form, the constraint model is

$$\text{maximize } Z_k(\boldsymbol{X}) \tag{10.13}$$

subject to:

$$g_i(\boldsymbol{X}) = b_i \quad i = 1, 2, \ldots, m \tag{10.14}$$

$$Z_j(\boldsymbol{X}) \geq L_j \quad \forall j \neq k. \tag{10.15}$$

Note that the dual variables associated with the right-hand-side values $L_j$ are the marginal rates of substitution or rate or change of $Z_k(\boldsymbol{X})$ per unit change in $L_j$ (or $Z_j(\boldsymbol{X})$ if binding).

Figure 10.14 illustrates the constraint method for a two-objective problem.

An efficiency frontier identifying the tradeoffs among conflicting objectives can be defined by solving the model many times for many values of the lower bounds. Just as with the weighting method, this can be a big job if there are many objectives. If there are more than two or three objectives, the results cannot be plotted. Pair-wise tradeoffs that can easily be plotted do not always clearly identify non-dominated alternatives.

The number of solutions to a weighting or constraint method model can be reduced considerably if the participants in the planning and management process can identify the acceptable weights or lower limits. However, this is not the language of decision-makers. Decision-makers who count on the support of each interest group are not content to assign weights that imply the relative importance of those various stakeholder interests. In addition, decision-makers should not be expected to know what they may want until they know what they can get, and at what cost (often politically as much as economically). However, there are ways of modifying the weighting or constraint methods to reduce the amount of effort in identifying these tradeoffs as well as the amount of information generated that is of no interest to those making decisions. This can be done using interactive methods that will be discussed shortly.

The weighting and constraint methods are among many methods available for generating efficient or non-inferior solutions (see, for example, Steuer, 1986). The use of methods that generate many solutions, even just efficient ones, assumes that once all the non-inferior alternatives have been identified, the participants in the planning and management process will be able to select the best compromise alternative from among them. In some situations this has worked. Undoubtedly, there will be planning activities in the future where the use of these non-inferior solution generation techniques alone will continue to be of value. However, in many other planning situations, they will not be sufficient in themselves. Often, the number of feasible non-inferior alternatives is simply too large. Participants in the planning and management process will not have the time or patience to examine and evaluate each alternative plan. Participants may also need help in identifying which alternatives they prefer. If they are willing to work with the analyst, the analyst can help them identify what alternatives they prefer without generating and comparing all the other non-inferior plans.

There are a number of methods available for assisting in selecting the most desirable non-dominated plan. Some of the more common ones are described next.

## 5.4. Satisficing

One method of further reducing the number of alternatives is called satisficing. It requires that the participants in the planning and management process specify a minimum acceptable value for each objective. Those alternatives that do not meet these minimum performance values are eliminated from further consideration. Those that remain can again be screened if the minimal acceptable values of one or more objectives are increased. When used in an iterative fashion, the number of non-inferior alternatives can be reduced down to a single best compromise plan or set of plans that everyone considers equally desirable. This process is illustrated in Figure 10.15.

Of course, sometimes the participants in the planning process will be unwilling or unable to narrow down the set of available non-inferior plans sufficiently with the iterative satisficing method. Then it may be necessary to examine in more detail the possible tradeoffs among the competing alternatives.

## 5.5. Lexicography

Another simple approach is called lexicography. To use this approach, the participants in the planning process must rank the objectives in order of priority. This ranking process takes place without considering their particular values. Then, from among the non-inferior plans that satisfy the minimum levels of each objective, the plan that is the best with respect to the highest priority objective will be the one selected as superior.

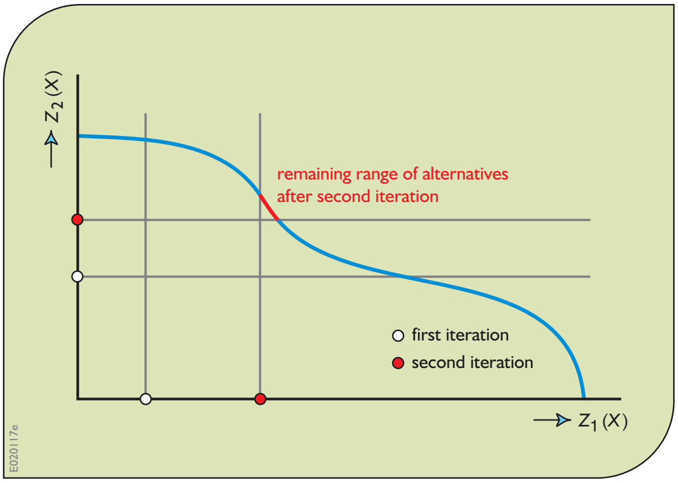



**Figure 10.15.** Two successive iterations of the satisficing method in which minimum levels of each objective are set, thereby eliminating those alternatives that do not meet these minimum levels.

If there is more than one plan that has the same value of the highest priority objective, then among this set of preferred plans the one that achieves the highest value of the second priority objective is selected. If here too there are multiple such plans, the process can continue until there is a unique plan selected.

This method assumes such a ranking of the objectives is possible. Often the relative values of the objectives of each alternative plan are of more importance to those involved in the decision-making process. Consider, for example, the problem of purchasing apples and oranges. Assuming you like both types of fruit, which should you buy if you have only enough money to buy one type? If you know you already have lots of apples, but no oranges, perhaps you would buy oranges, and vice versa. Hence, the ranking of objectives can depend on the current state and needs of those who will benefit from the plan.

## 5.6. Indifference Analysis

Another method of selecting the best plan is called indifference analysis. To illustrate the possible application of indifference analysis to plan selection, consider a simple situation in which there are only two alternative plans ($A$ and $B$) and two planning objectives (1 and 2) being considered. Let $Z_1^A$ and $Z_2^A$ be the values of the two respective objectives for plan $A$ and $Z_1^B$ and $Z_2^B$ be the values of the two respective objectives for plan $B$. Comparing both plans when one objective is better than another for each plan can be difficult. Indifference analysis can reduce the problem to one of comparing the values of only one objective.

Indifference analysis first requires the selection of an arbitrary value for one of the objectives, say $Z_2^*$, for objective 2. It is usually a value within the range of the values $Z_2^A$ and $Z_2^B$, or in a more general case between the maximum and minimum of all objective 2 values.

Next, a value of objective 1, say $Z_1$, must be selected such that the participants involved are indifferent between the hypothetical plan that would have as its objective values $(Z_1, Z_2^*)$ and plan $A$ that has as its objective values $(Z_1^A, Z_2^A)$. In other words, $Z_1$ must be determined such that $(Z_1, Z_2^*)$ is as desirable as or equivalent to $(Z_1^A, Z_2^A)$:

$$(Z_1, Z_2^*) \approx (Z_1^A, Z_2^A) \tag{10.16}$$

Next another value of the first objective, say $Z_1'$, must be selected such that the participants are indifferent between a hypothetical plan $(Z_1', Z_2^*)$ and the objective values $(Z_1^B, Z_2^B)$ of plan $B$:

$$(Z_1', Z_2^*) \approx (Z_1^B, Z_2^B) \tag{10.17}$$

These comparisons yield hypothetical but equally desirable plans for each actual plan. These hypothetical plans differ only in the value of objective 1 and, hence, they are easily compared. If both objectives are to be maximized and $Z_1$ is larger than $Z_1'$, then the first hypothetical plan yielding $Z_1$ is preferred to the second hypothetical plan yielding $Z_1'$. Since the two hypothetical plans are equivalent to plans $A$ and $B$, respectively, plan $A$ must be preferred to plan $B$. Conversely, if $Z_1'$ is larger than $Z_1$, then plan $B$ is preferred to plan $A$.

This process can be extended to a larger number of objectives and plans, all of which may be ranked by a common objective. For example, assume that there are three objectives $Z_1^i$, $Z_2^i$, $Z_3^i$, and $n$ alternative plans $i$. A reference value $Z_3^*$ for objective 3 can be chosen and a value $\underline{Z}_1^i$ estimated for each alternative plan $i$ such that one is indifferent between $(\underline{Z}_1^i, Z_2^i, Z_3^*)$ and $(Z_1^i, Z_2^i, Z_3^i)$. The second objective value remains the same as in the actual alternative in each of the hypothetical alternatives. Thus, the focus is on the tradeoff between the values of objectives 1 and 3. Assuming that each objective is to be maximized, if $Z_3^*$ is selected so that $Z_3^* < Z_3^i$, *then* $\underline{Z}_1^i$ *will no doubt be greater than* $Z_3^i$. *Conversely, if* $Z_3^* > Z_3^i$, *then* $\underline{Z}_1^i$ *will be less than* $Z_3^i$.

Next, a new hypothetical plan containing a reference value $Z_2^*$ and $Z_3^*$ can be created and compared with $(\underline{?}, Z_2^i, Z_3^*)$. The focus is on the tradeoff between the values of objectives 1 and 2. A value of $Z_1^{i\prime}$ must be selected such that the participants are indifferent between $(Z_1^{i\prime}, \underline{Z}_2^i, Z_3^*)$ and $(\underline{Z}_1^i, Z_2^*, Z_3^*)$. Hence, for all plans $i$, the participants are indifferent between two hypothetical plans and the actual one. The last hypothetical plans differ only by the value of the first objective. The plan that has the largest value for objective 1 will be the best plan. This was achieved by pair-wise comparisons only.

In the first step objective 2 remained constant and only objectives 1 and 3 were compared to get

$$(Z_1^i, Z_2^i, Z_3^i) \approx (\underline{Z}_1^i, Z_2^i, Z_3^*) \tag{10.18}$$

In the next step involving the hypothetical plans just defined, objective 3 remained constant and only objectives 1 and 2 were compared to get

$$(\underline{Z}_1^i, Z_2^i, Z_3^*) \approx (Z_1^{i\prime}, Z_2^*, Z_3^*) \tag{10.19}$$

Hence

$$(Z_1^i, Z_2^i, Z_3^i) \approx (\underline{Z}_1^i, Z_2^i, Z_3^*) \approx (Z_1^{i\prime}, Z_2^*, Z_3^*) \tag{10.20}$$

Having done this for all $n$ plans, there are now $n$ hypothetical plans $(Z_1^{i\prime}, Z_2^*, Z_3^*)$ that differ only in the value of $Z_1^{i\prime}$. All $n$ plans can be ranked just on the basis of the value of this single objective.

Each of these plan selection techniques requires the prior identification of discrete alternative plans.

## 5.7. Goal Attainment

The goal attainment method combines some of the advantages of both the weighting and constraint plan generation methods already discussed. The participants in the planning and management process specify a set of goals or targets $T_j$ for each objective $j$ and, if applicable, a weight, $w_j$, that reflects the relative importance of meeting that goal compared to meeting other goals. If the participants are unable to specify these weights, the analyst must select them and then later change them on the basis of their reactions to the generated plans.

The goal attainment method identifies the plans that minimize the maximum weighted deviation of any

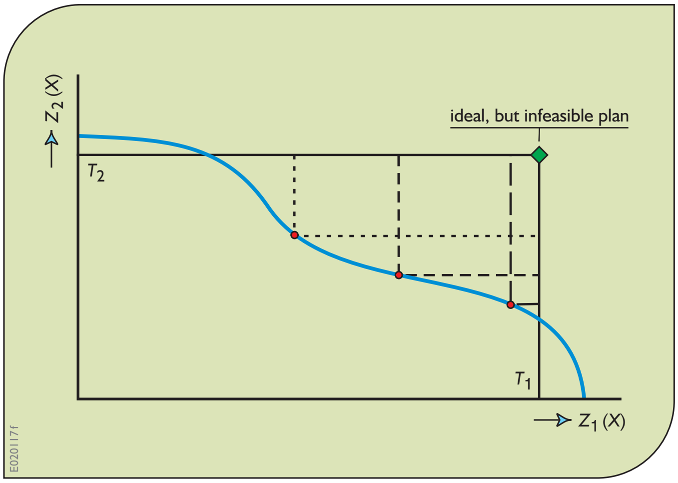



**Figure 10.16. The goal attainment method of generating points on the efficiency frontier using different values of the weights $w_1$ and $w_2$ for fixed objective target values $T_1$ and $T_2$.**

objective value, $Z_j(\boldsymbol{X})$, from its specified target, $T_j$. The problem is to

$$\text{minimize D} \tag{10.21}$$

subject to:

$$g_i(\boldsymbol{X}) = b_i \qquad i = 1, 2, \dots, m \tag{10.22}$$

$$w_j[T_j - Z_j(\boldsymbol{X})] \leq D \quad j = 1, 2, \dots, J \tag{10.23}$$

Constraints 10.22 contain the relationships among the decision-variables in the vector $\boldsymbol{X}$. They define the feasible region of decision-variable values.

This method of multi-criteria analysis can generate efficient or non-inferior plans by adjusting the weights and targets. It is illustrated for a two-objective problem in Figure 10.16.

Unless $T_j \geq Z_j(\boldsymbol{X})$, some plans generated from goal programming may be inferior with respect to the objectives being considered.

## 5.8. Goal-Programming

Goal-programming methods also require specified target values, along with relative losses or penalties associated with deviations from these target values. The objective is to find the plan that minimizes the sum of all such losses or penalties. Assuming for this illustration that all such losses can be expressed as functions of deviations from target values, and again assuming each objective is to be maximized, the general goal-programming problem is to

$$\text{minimize } \sum_j v_j\, D_j + w_j\, E_j \tag{10.24}$$

subject to:

$$g_i(\boldsymbol{X}) = b_i \qquad i = 1, 2, \dots, m \tag{10.25}$$

$$Z_j(\boldsymbol{X}) = T_j^* - D_j + E_j \tag{10.26}$$

where the parameters $v_j$ and $w_j$ are the penalties (weights) assigned to objective value deficits or excesses, as appropriate. The weights and the target values, $T_j^*$, can be changed to get alternative solutions, or tradeoffs, among the different objectives.

## 5.9. Interactive Methods

Interactive methods allow participants in the planning process to explore the range of possible decisions without having to generate all of them, especially those of little interest to the participants.

One such approach, called the step method, is an iterative method requiring preference information from the participants in the planning and management process at each iteration. This information identifies constraints on various objective values. The weighting method is used to get an initial solution on the efficiency frontier. The weights, $w_j$, are calculated on the basis of the relative range of values each objective $j$ can assume, and on whether or not the participants in the planning and management process have indicated satisfaction regarding a particular objective value obtained from a previous solution. If they are satisfied with the value of, say, an objective $Z_j(\boldsymbol{X})$, they must indicate how much of that value they would be willing to give up to obtain unspecified increases in objectives whose values they consider unsatisfactory. This defines a lower bound on $Z_j(\boldsymbol{X})$. Then the weight $w_j$ for that objective is set to 0, and the weights of all remaining objectives are re-calculated. The problem is again solved. The process is repeated until some best compromise plan is identified.

This step method guides the participants in the planning and management process among non-inferior alternatives toward the plan or solution the participants consider best without requiring an exhaustive generation of all non-inferior alternatives. Even if the best compromise solution is not identified or agreed upon, the method provides a way for participants to learn what the tradeoffs are in the region of solutions of interest to them. However, the participants must be willing to indicate how much of some objective value can be reduced to obtain some unknown improvement of other objective values. This is not as easy as indicating how much more is desired of any or all objectives whose values are unsatisfactory.

To overcome this objection to the step method, other interactive methods have been developed. These begin with an obviously inferior solution. Based on a series of questions concerning how much more important it is to obtain various improvements of each objective, the methods proceed from that inferior solution to more improved solutions. The end result is either a solution everyone agrees is best, or an efficient one where no more improvements can be made in one objective without decreasing the value of another.

These iterative interactive approaches are illustrated in Figure 10.17. To work, they require the participation of the participants in the planning and management process.

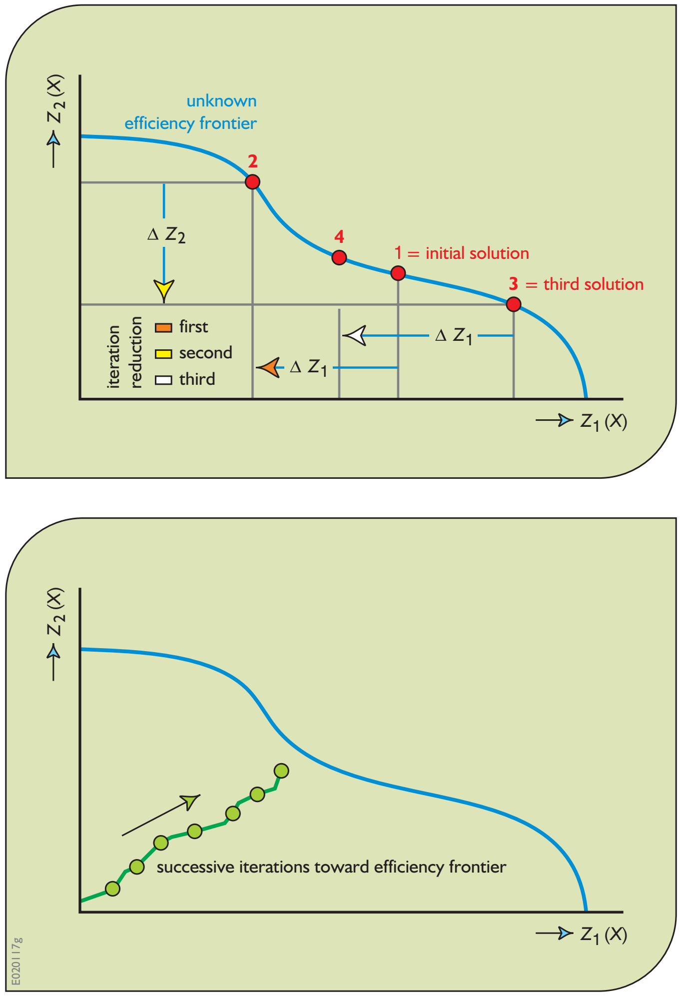



**Figure 10.17. Two interactive iterative multiple criteria approaches for identifying the tradeoffs of interest and possibly the best decision.**

## 5.10. Plan Simulation and Evaluation

The methods outlined above provide a brief introduction to some of the simpler approaches available for plan identification and selection. Details on these and other potentially useful techniques can be found in many books, some of which are devoted solely to this subject of multi-objective planning (Cohon, 1978; Steuer, 1986). Most have been described in an optimization framework. This section describes ways of evaluating alternative water management plans or policies on the basis of the time series of performance criteria values derived from simulation models.

Simulation models of water resources systems yield time series of hydrological variable values. These values in turn affect multiple system performance criteria, each pertaining to a specific interest and measured in its appropriate units. A process for evaluating alternative water resources management plans or policies based on these simulation model results includes the following steps:

1. Identify system performance indicators that are affected by one or more hydrological attributes whose values will vary depending on the management policy or plan being simulated. For example, navigation benefits, measured in monetary units, might depend on water depths and velocities. Hydropower production is affected by lake levels and discharges through the power plant. Water quality might be expressed as the average or maximum concentration over a fixed time period at certain locations for various potential pollutants, and will depend in part on the flows. Ecological habitats may be affected by flows, water depths, water quality, flooding frequency or duration, and/or rates of changes in these attributes.
2. Define the functional relationships between these performance indicators and the hydrological attributes. Figure 10.18 illustrates such functions. The units on each axis may differ for each such function.
3. Simulate to obtain time series of hydrological attribute values and map them into a time series of performance

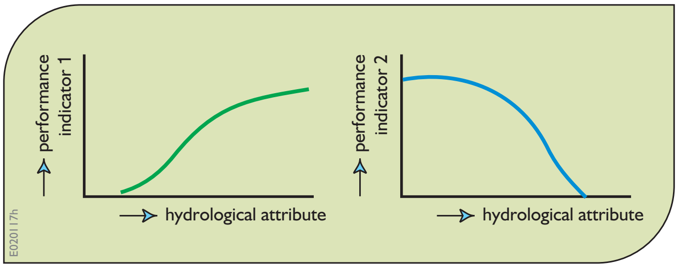



**Figure 10.18.** Performance indicators expressed as functions of simulated hydrological attributes.

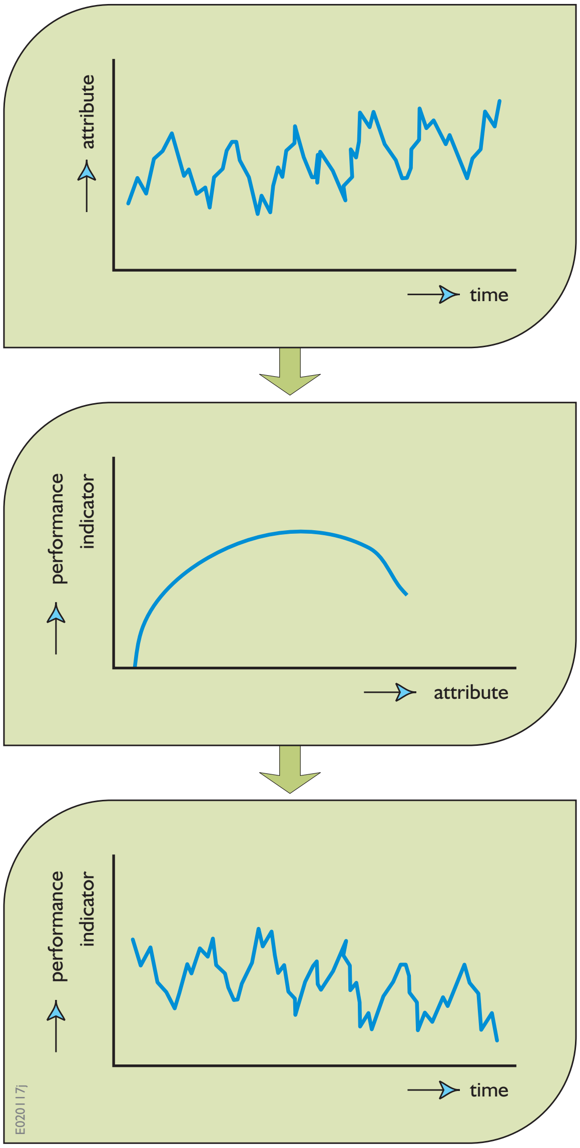



**Figure 10.19.** Mapping a time series of hydrological attribute values into a corresponding time series of performance indicator values.

indicator values using the functional relationships defined in Step 2. This step is illustrated in Figure 10.19.

4. Combine multiple time-series values for the same performance criterion, as applicable, as shown in Figure 10.20. This can be done using maximum or minimum values, or arithmetic or geometric means, as appropriate. For example, flow velocities, depths and algal biomass concentrations may affect recreational boating. The three sets of time series of recreational boating benefits or suitability can be combined into one time series, and statistics of this overall time series can be compared to similar statistics of other performance indicators. This step gives the modeller an opportunity to calibrate the resulting single system performance indicator.
5. Develop and compare system performance exceedance distributions, or divide the range of performance values into colour-coded ranges and display on maps or in scorecards, as illustrated in Figure 10.21.

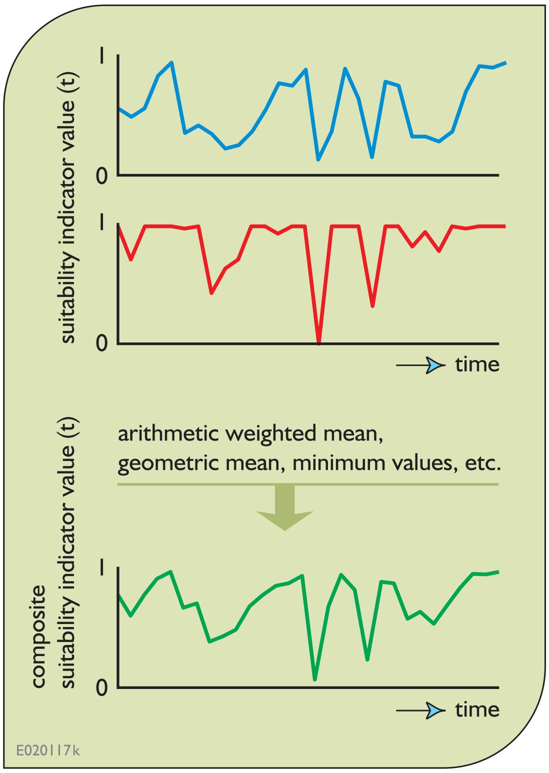
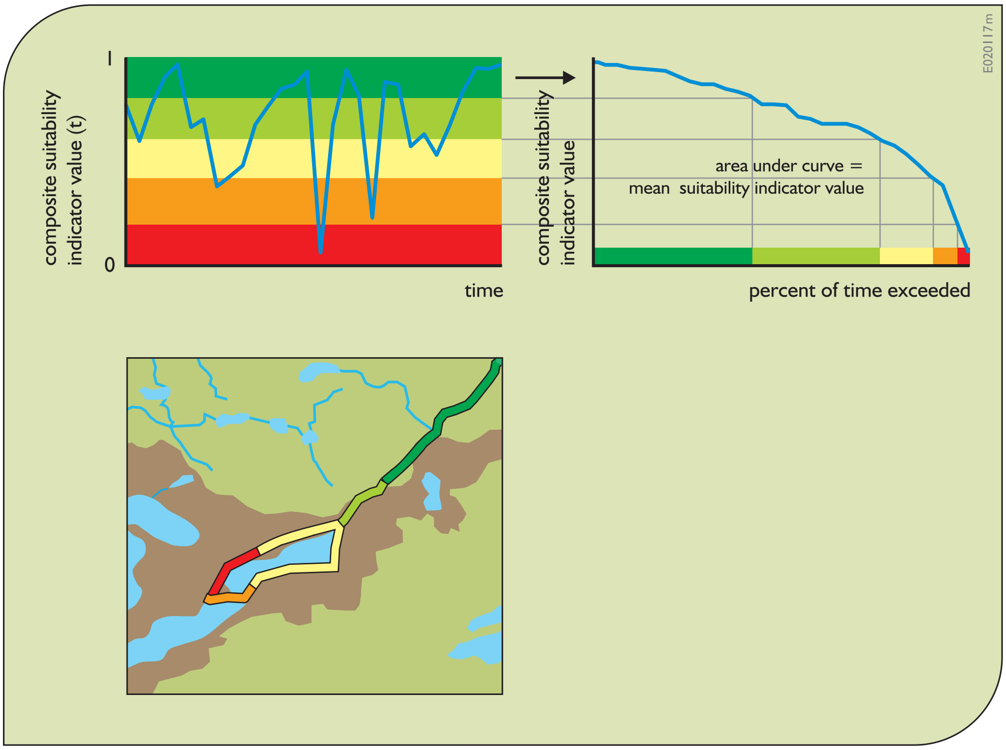



**Figure 10.20.** Combining multiple time series of values of a specified performance indictor into a single time series of values of that performance indicator.




E020117o

alternative strategy

| performance indicators | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 5 | 1 | 6 | 8 | 3 | 2 |
| 2 | 10 | 12 | 65 | 4 | 57 | 95 | 34 |
| 3 | – | – – | + + + | – – – | | + | – – |
| 4 | 0.2 | 0.8 | 0.5 | 1.0 | 0.0 | 0.7 | 0.4 |
| 5 | B | C | D | B | A | B | C |
| 6 | 100 | 45 | 49 | 69 | 78 | 34 | 22 |

**Figure 10.21.** Ways of summarizing and displaying time-series performance indicator data involving exceedance distributions, and colour-coded maps and scorecards. Colour-coded map displays on computers can be dynamic, showing changes over time. Green and red coloured scorecard entries indicate best and worst plan or strategy, respectively, for associated performance indicator (for scorecards see also Appendix E Section 6.4).

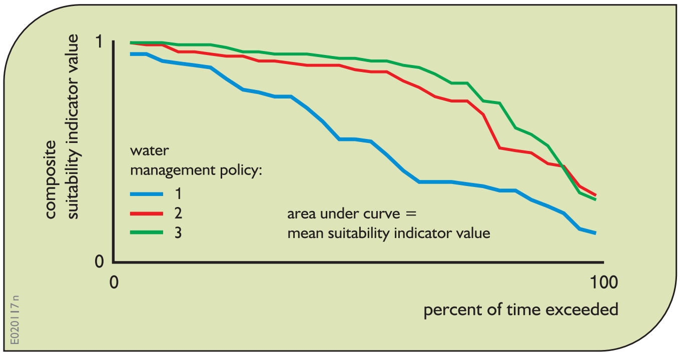



**Figure 10.22.** Three exceedance functions showing decrease in mean performance value if Policy 1 is followed compared to Policies 2 and 3.

The area under each exceedance curve is the mean. Different exceedance functions will result from different water management policies, as illustrated in Figure 10.22.

One can establish thresholds to identify discrete zones of performance indicator values, and assign a colour to each zone. Measures of reliability, resilience and vulnerability can then be calculated and displayed as well.

Scorecards can show the mean values of each indicator for any set of sites. The best value for each indicator can be coloured green; the worst value for each indicator can be coloured red. The water management alternative having the most green boxes will stand out and will probably be considered more seriously than the alternative having the highest number of red (worst value) boxes.

This five-step process has been used in a study of improved ways of managing lake levels and flows in Lake Ontario and its discharges into the St. Lawrence River in North America. Performance criteria were defined for domestic and industrial water supplies, navigation depths, shore bank erosion, recreational boating, hydropower production, flooding, water quality and ecological habitats. The performance measures for each of these interests were identified and expressed as functions of one or more hydrological variable values related to flow and lake-level management. Models designed to simulate alternative lake-level and flow-management policies were used to generate sets of time series for each system performance criterion. These in turn were combined, summarized and compared.

The same five-step process has been implemented in the Everglades restoration project in southern Florida. The Everglades is a long, very wide and extremely flat 'river of grass' flowing generally south into, eventually, the Atlantic Ocean and Gulf of Mexico. This ecosystem restoration project has involved numerous local, state and federal agencies. The project affects a large population and agricultural industry that want secure and reliable water supplies and flood protection. Its current estimated cost over some three decades is about $8 billion. Hence, it involves politics. But its goal is primarily focused on restoring a unique ecosystem that is increasingly degraded due to extensive alterations in its hydrology over the past half-century.

The motto of the Everglades restoration project is 'to get the water right.' Those who manage the region's water are attempting to restore the ecosystem by restoring the hydrological regime – the flows, depths, hydroperiods and water qualities – throughout the region to what they think existed some fifty years ago. The trick is to accomplish this and still meet the water supply, flood protection and land development needs of those who live in the region. Clearly, achieving a hydrological condition that existed before people began populating that region in significant numbers will not be possible. Hence the question: what if water managers are not able to 'get the water right'? What if they can only get the water right on 90% of the area, or what if they can only get 90% of the water right on all the area? What will be the likely impact on the ecosystem? Are there opportunities for changing hydrology to improve ecology? Where?

To address questions such as these, at least in a preliminary way until more detailed ecological models become available, the five-step approach outlined above is being applied in the Everglades. It is being used to extend simulated hydrological predictions to produce relative values of ecological habitat suitability indicators for selected indicator species, as illustrated in Figure 10.8, and topographic characteristics.

This five-step procedure does not find an 'optimal' water management policy. It can however contribute useful information to the political debate that must take place in the search for that optimum. Each step of the approach can and should include and involve the various stakeholders and publics in the basin. These individuals are sources of important inputs in this evaluation process. Stakeholders who will be influencing or making water-management decisions need to

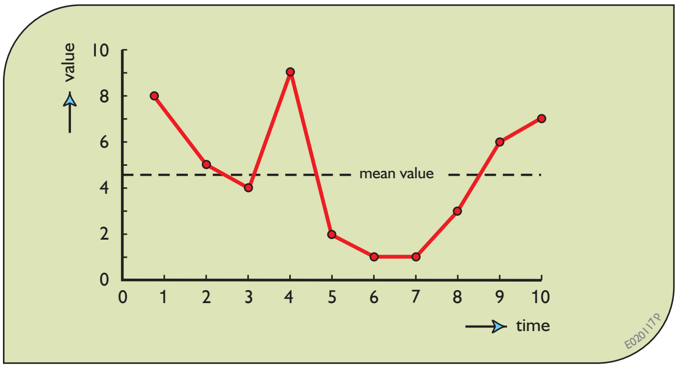



Figure 10.23. Plot of time-series data having a mean value of 4.6 and a variance of 7.44.

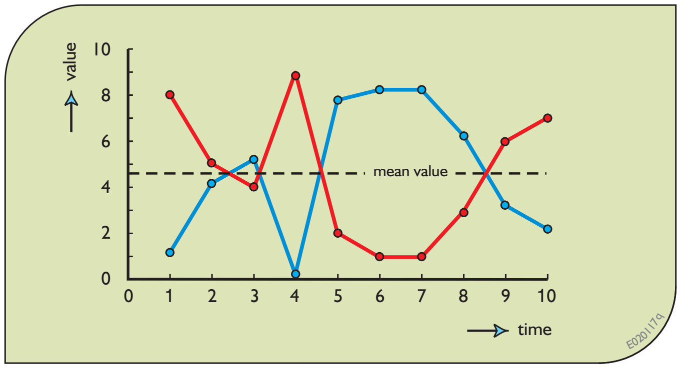



Figure 10.24. A plot of two different time series having the same mean and variance.

understand just how this multi-objective evaluation process works if they are to accept and benefit from its results. Stakeholder involvement in this process can help lead to a common understanding (or 'shared vision') of how their system works and of the tradeoffs that exist among conflicting objectives. The extent to which all stakeholders understand this evaluation approach or procedure and how it is applied in their basin will largely determine their ability to participate effectively in the political process of selecting the best water management policy or practice.

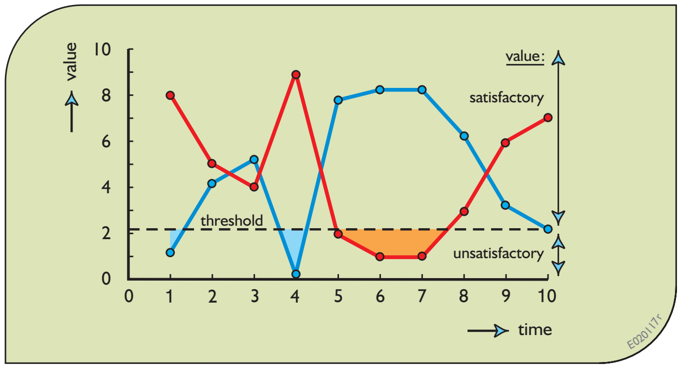



Figure 10.25. Threshold value distinguishing values considered satisfactory, and those considered unsatisfactory.

## 6. Statistical Summaries of Performance Criteria

There are numerous ways of summarizing time-series performance data that might result from simulation analyses. Weighted arithmetic mean values or geometric mean values are two ways of summarizing multiple time-series data. The overall mean itself generally provides too little information about a dynamic process. Multiple time-series plots themselves are often hard to compare.

Another way to summarize and compare time-series data is to calculate and compare the variance of the data.

Consider a time series of $T$ values $X_t$ whose mean is $\underline{X}$. For example, suppose the time series consisted of 8, 5, 4, 9, 2, 1, 1, 3, 6, and 7. The mean of these 10 values is 4.6. The variance is

$$\sum_t^T (X_t - \underline{X})^2/T = [(8-4.6)^2 + (5-4.6)^2 + \cdots + (7-4.6)^2]/10 = 7.44 \qquad (10.27)$$

A plot of these values and their mean is shown in Figure 10.23.

The mean and variance for the time series shown in Figure 10.23, however, are the same for its upside-down image, as shown in Figure 10.24. They do not even depend on the order of the time-series data.

Consider these two sets of time series shown again in Figure 10.25, each having the same mean and variance. Assume that any value equal to or less than the dashed line (just above 2) is considered unsatisfactory. This value is called a *threshold value*, dividing the time-series data into satisfactory and unsatisfactory values.

It is clear from Figure 10.25 that the impact of these two time series could differ. The original time series shown in a red line remained in an unsatisfactory condition for a longer time than did the 'rotated' time series shown in blue. However, its maximum extent of failure when it failed was less than the rotated time series. These characteristics can be captured by the measures of reliability, resilience and vulnerability (Hashimoto et al., 1982).

### 6.1. Reliability

The reliability of any time series can be defined as the number of data in a satisfactory state divided by the total number of data in the time series. Assuming satisfactory values in the time series $X_t$ containing $n$ values are those equal to or greater than some threshold $X^T$, then

$$\text{Reliability}[\boldsymbol{X}] = [\text{number of time periods } t \text{ such that } X_t \geq X^T]/n \quad (10.28)$$

The reliability of the original time series shown in red is 0.7. It failed three times in ten. The reliability of the rotated time series shown in blue is also 0.7, failing three times in ten.

Is a more reliable system better than a less reliable system? Not necessarily. Reliability measures tell one nothing about how quickly a system recovers and returns to a satisfactory value, nor does it indicate how bad an unsatisfactory value might be should one occur. It may well be that a system that fails relatively often, but by insignificant amounts and for short durations, will be much preferable to one whose reliability is much higher but where, when a failure does occur, it is likely to be much more severe. Resilience and vulnerability measures can quantify these system characteristics.

### 6.2. Resilience

Resilience can be expressed as the probability that if a system is in an unsatisfactory state, the next state will be satisfactory. It is the probability of having a satisfactory value in time period $t + 1$, given an unsatisfactory value in any time period $t$. It can be calculated as

$$\text{Resilience}[\boldsymbol{X}] = [\text{number of times a satisfactory value follows an unsatisfactory value}]/[\text{number of times an unsatisfactory value occurred}] \quad (10.29)$$

Resilience is not defined if no unsatisfactory values occur in the time series. For the original time series shown in red, the resilience is 1/3, again assuming the value of 2 or less is considered a failure. For the rotated time series shown in blue the resilience is $2/2 = 1$. We cannot judge the resilience of the blue time series on the basis of the last failure in period 10 because we do not have an observation in period 11.

### 6.3. Vulnerability

Vulnerability is a measure of the extent of the differences between the threshold value and the unsatisfactory time-series values. Clearly, this is a probabilistic measure. Some use expected values, some use maximum observed values, and others may assign a probability of exceedance to their vulnerability measures. Assuming an expected value measure of vulnerability is to be used:

$$\text{Vulnerability}[\boldsymbol{X}] = [\text{sum of positive values of } (X^T - X_t)]/[\text{number of times an unsatisfactory value occurred}] \quad (10.30)$$

The expected vulnerability of the original red time series is $[(2 - 2) + (2 - 1) + (2 - 1)]/3 = 0.67$. The expected vulnerability of the time series shown by the rotated blue line on Figure 10.25 is $[(2 - 1.2) + (2 - 0.2)]/2 = 1.3$.

So, depending on whether a threshold value is considered a failure or not in this example, the reliability and resilience of original (red) time series is equal or less than the rotated (blue) time series. However, the expected vulnerability of the original red time series is less than that of the rotated blue time series. This shows the typical tradeoffs one can define using these three measures of system performance.

## 7. Conclusions

Many theoretical and practical approaches have been proposed in the literature for identifying and quantifying objectives and for considering multiple criteria or objectives in water resources planning. The discussion and techniques presented in this chapter serve merely as an introduction to this subject. These tools, including their modifications and extensions, are designed to provide information that can be of value to the planning and decision-making process.

Water resources systems planners and managers and the numerous other participants typically involved in decision-making face a challenge when they are required to select one of many alternatives, each characterized by different values among multiple performance criteria.

It requires a balancing of the goals and values of the various individuals and groups concerned with the project. There is virtually no way in which the plan-selection step can be a normative process or procedure; there can be no standard set of criteria or methods which will identify the preferred project. At best, an iterative procedure in which those using tools similar to those described in this chapter together with all the interested stakeholders may reach some shared vision of what is best to do, at least until conditions change or new knowledge or new goals or new requirements emerge. This may be the only way to identify a plan that is politically as well as technically, socially, financially and institutionally feasible.

To many participants in the planning process, some of these approaches for objective quantification and multi-objective planning may seem theoretical or academic. Many may be reluctant to learn quantitative policy analysis techniques or to spend time answering seemingly irrelevant questions that might lead eventually to a 'compromise' plan. Reluctance to engage in quantification of tradeoffs among particular objectives of alternative plans, and by implication alternative stakeholder groups, may stem from the support decision-makers desire from all of these conflicting interest groups. In such situations it is obviously to their advantage not to be too explicit in quantifying political values. They might prefer the 'analyst' to make these tradeoffs so that they just need not discuss them. (We writers have participated in such situations.) Planners, engineers or analysts are often very willing to make these tradeoffs because they pertain to subject areas in which they often consider themselves experts. When political tradeoffs are at issue, no one is an expert. No one has the 'optimal' answer, but professionals should be engaged in and informing and facilitating the process of coming to an acceptable, and often compromise, decision.

Through the further development and use of practical analytical multi-objective planning techniques, analysts can begin to interact with all participants in the planning and management process and can enlighten any who would argue that water resources policy evaluation and analyses should not be political. Analysts, managers and planners have to work in a political environment. They need to understand the process of decision-making, what information is most useful to that process, and how it can best be presented. Knowledge of these facts in a particular planning situation might largely dictate the particular approach to objective identification and quantification and to plan selection that is most appropriate.

The method deemed most appropriate for a particular situation will depend not only on the physical scale of the situation itself but also on the decision-makers, the decision-making process, and the responsibilities accepted by the analysts, the participants and the decision-makers.

Finally, remember that the decisions being made at the current time are only part of a sequence of decisions that will be made on into the future. No one can predict with certainty what future generations will consider as being important or what they will want to do, but spending some time trying to guess is not an idle exercise. It pays to plan ahead, as best one can, and ask ourselves if the decisions being considered today will be those we think our descendants would have wanted us to make. This kind of thinking gets us into issues of adaptive management (Appendix B) and sustainability (ASCE, 1999).

## 8. References

ASCE (American Society of Civil Engineers). 1999. *Sustainability criteria for water resource systems*. Reston, Va., ASCE Press, (also D.P. Loucks and J.S. Gladwell (eds.), Cambridge, U K, Cambridge University Press, 2000).

COHON, J.L. 1978. *Multiobjective programming and planning*. New York, Academic Press.

GREGORY, R.S. and KEENEY, R.L. 1994. *Creating policy alternatives using stakeholder values*. *Management Science*, Vol. 40, pp. 1035–48.

HAMMOND, J.S.; KEENEY, R.L. and RAIFFA, H. 1999. *Smart choices: a practical guide to making better decisions*. Boston, Mass., Harvard Business School Press.

HASHIMOTO, T.; LOUCKS, D.P. and STEDINGER, J.R. 1982. Robustness of water resource systems. *Water Resources Research*, Vol. 18, No. 1, pp. 21–26.

HASHIMOTO, T.; STEDINGER, J.R. and LOUCKS, D.P. 1982. Reliability, resiliency and vulnerability criteria for water resource system performance evaluation. *Water Resources Research*, Vol. 18, No. 1, pp. 14–20.

STEUER, R.E. 1986. *Multiple criteria optimization: theory, computation, and application*. New York, Wiley.

## Additional References (Further Reading)


CHEN, H.W. and CHANG, N.B. 1998. Water pollution control in the river basin by fuzzy genetic algorithm-based multiobjective programming modelling. *Water Science and Technology*, Vol. 37, No. 8, pp. 55–63.

CIENIAWSKI, S.E.; EHEART, J.W. and RANJITHAN, S. 1995. Using genetic algorithms to solve a multiobjective groundwater monitoring problem. *Water Resources Research*, Vol. 31, No. 2, February, pp. 399–409.

ESCHENAUER, H.; KOSKI, J. and OSYCZKA, A. 1990. *Multicriteria design optimization*. Berlin, Springer-Verlag.

GUPTA, H.V.; SORROSHIAN, S. and YAPO, P.O. 1998. Towards improved calibration of hydrological models: multiple and non-commensurable measures of information. *Water Resources Research*, Vol. 34, No. 4, pp. 751–63.

HAIMES, Y.Y.; HALL, W.A. and FREEDMAN, H.T. 1975. *Multiobjective optimization in water resources systems*. Amsterdam, Elsevier Scientific.

HALHAL, D.; WALTERS, G.A.; OUAZAR, D. and SAVIC, D.A. 1997. Multiobjective improvement of water distribution systems using a structure messy genetic algorithm approach. *Journal of Water Resources Planning and Management ASCE*, Vol. 123, No. 3, pp. 137–46.

KEENEY, R.L. and RAIFFA, H. 1976. *Decisions with multiple objectives: preferences and value tradeoffs*. New York, Wiley.

LOUCKS, D.P.; STEDINGER, J.R. and HAITH, D.A. 1981. *Water resource systems planning and analysis*. Englewood Cliffs, N.J., Prentice Hall.

MAASS, A.; HUFSCHMIDT, M.M.; DORFMAN, R.; THOMAS, H.A. Jr.; MARGLIN, S.A. and FAIR, G.M. 1962. *Design of water resource systems*. Cambridge, Mass., Harvard University Press.

REED, P.M.; MINSKER, B.S. and GOLDBERG, D.E. 2001. A multiobjective approach to cost effective long-term groundwater monitoring using an elitist nondominated sorted genetic algorithm with historical data. *Journal of Hydroinformatics*, Vol. 3, No. 2, April, pp. 71–89.

RITZEL, B.J.; EHEART J.W. and RANJITHAN, S. 1994. Using genetic algorithms to solve a multiple objective groundwater pollution containment problem. *Water Resources Research*, Vol. 30, No. 5, May, pp. 1589–1603.

SAWARAGI, Y.; NAKAYAMA, H. and TANINO, T. 1985. *Theory of multiobjective optimization*. Orlando, Fla., Academic Press.

STATNIKOV, R.B. and MATUSOV, J.B. 1995. *Multicriteria optimization and engineering*. New York, Chapman and Hall.

YEH, C.H. and LABADIE, J.W. 1997. Multiobjective Watershed-level planning of storm-water detention systems. *Journal of Water Resources Planning and Management*, Vol. 123, No. 6, November/December, pp. 3360–43.

ZELENY, M. 1982. Multiple criteria decision making. New York, McGraw-Hill.

# 11. River Basin Planning Models

1. Introduction 325
   1.1. Scales of River Basin Processes 326
   1.2. Model Time Periods 327
   1.3. Modelling Approaches For River Basin Management 328

2. Modelling the Natural Resource System and Related Infrastructure 328
   2.1. Watershed Hydrological Models 328
      2.1.1. Classification of Hydrological Models 329
      2.1.2. Hydrological Processes: Surface Water 329
      2.1.3. Hydrological Processes: Groundwater 333
      2.1.4. Modelling Groundwater: Surface Water Interactions 336
      2.1.5. Streamflow Estimation 339
      2.1.6. Streamflow Routing 341
   2.2. Lakes and Reservoirs 342
      2.2.1. Estimating Active Storage Capacity 343
      2.2.2. Reservoir Storage–Yield Functions 344
      2.2.3. Evaporation Losses 346
      2.2.4. Over and Within-Year Reservoir Storage and Yields 347
      2.2.5. Estimation of Active Reservoir Storage Capacities for Specified Yields 348
   2.3. Wetlands and Swamps 354
   2.4. Water Quality and Ecology 354

3. Modelling the Socio-Economic Functions In a River Basin 355
   3.1. Withdrawals and Diversions 355
   3.2. Domestic, Municipal and Industrial Water Demand 356
   3.3. Agricultural Water Demand 357
   3.4. Hydroelectric Power Production 357
   3.5. Flood Risk Reduction 359
      3.5.1. Reservoir Flood Storage Capacity 360
      3.5.2. Channel Capacity 362
   3.6. Lake-Based Recreation 362

4. River Basin Analysis 363
   4.1. Model Synthesis 363
   4.2. Modelling Approach Using Optimzation 364
   4.3. Modelling Approach Using Simulation 365
   4.4. Optimization and/or Simulation 368
   4.5. Project Scheduling 368

5. Conclusions 371

6. References 371

# 11 River Basin Planning Models

Multipurpose river basin management typically involves the identification and use of both structural and non-structural measures designed to increase the reliability of municipal, industrial and agriculture water supplies when and where demanded, to protect against floods, to improve water quality, to provide for commercial navigation and recreation, and to produce hydropower, as appropriate for the particular river basin. Structural measures may include diversion canals, reservoirs, hydropower plants, levees, flood proofing, irrigation delivery and drainage systems, navigation locks, recreational facilities, groundwater wells, and water and wastewater treatment plants along with their distribution and collection systems. Non-structural measures may include land use controls and zoning, flood warning and evacuation measures, and economic incentives that affect human behaviour with regard to water and watershed use. Planning the development and management of water resources systems involves identifying just what, when and where structural or non-structural measures are needed, the extent to which they are needed, and their combined economic, environmental, ecological and social impacts. This chapter introduces some modelling approaches for doing this. The focus is on water quantity management. The following chapter reviews some measures and models for water quality management.

## 1. Introduction

This chapter introduces some types of models commonly used to assist those responsible for planning and managing various components of river systems. These components include the watersheds that drain into the surface water bodies and underlying aquifers of river systems. They include the streams, rivers, lakes, reservoirs and wetlands that can exist in river basins and that are affected by water management policies and practices. First, each of these components will be examined and modelled separately. The management of any single component, however, can affect the performance of other components in a river basin system. Hence, for the overall management of river basin systems, a systems view is needed. Typically, this systems view requires the modelling of multiple components. These multi-component models can then be used to analyse alternative designs and management strategies for integrated multi-component systems.

River basin planning is a prerequisite for integrated water resources management (IWRM). IWRM requires the integration of the natural system components (surface water–groundwater, quantity–quality, land- and water management, etc.) and the upstream and downstream water-related demands or interests. Water resources planning is increasingly done on a river basin scale. The European Water Framework Directive, for example, imposes the development of basin plans in Europe, forcing riparian countries to work together on the development and management of their river basins.

The discussion in this chapter is limited to water quantity management. Clearly the regimes of flows, velocities, volumes and other properties of water quantity will affect the quality of that water as well. Moreover, from the perspective of IWRM, quantity and quality aspects should be considered jointly. However, unless water quantity management strategies are based on requirements for water quality, such as for the dilution of pollutants, water quality measures do not normally affect water quantity. For this reason it is common to separate analytical approaches of water quantity management from those of water quality management. However, when attempting to predict the impacts of any management policy on both water quantity and quality, both quantity and quality models are needed.

This chapter begins with an introduction on the scales of river basin processes and an overview of modelling approaches. Section 2 describes the modelling of the natural resources system (see also Figure 1.19 in Chapter 1). This includes the estimation of unregulated surface and groundwater runoff from watersheds, as well as the prediction of streamflows at various sites of interest throughout a basin. This section also includes several methods for estimating reservoir storage requirements for water supplies and flood storage. Reservoirs can provide for water supply, flood control, recreation and hydroelectric power generation. These 'use-functions' and their modelling are described in Section 3 on socio-economic functions. The natural resource system and the socio-economic functions are then, in Section 4, combined into a multiple-purpose multi-objective planning model for a river basin. The chapter concludes with an introduction to some dynamic models for assisting in the scheduling and time sequencing of multiple projects within a river basin.

### 1.1. Scales of River Basin Processes

One of the challenges in constructing integrated models of multiple river basin processes is the wide range of spatial and temporal scales that characterize these processes. Precipitation, which is the source of virtually all freshwater in the hydrological cycle, is typically highly variable and uneven in its distribution over time and space. Similarly, the rates of evaporation and transpiration vary considerably according to climatic and land-cover conditions. The relative magnitudes of the fluxes associated with individual components of the hydrological cycle, such as evapotranspiration, may differ significantly even at small spatial scales such as on a commercial shopping mall, an agricultural field and a woodland. Surface-flow conditions that are of interest to those managing water supplies, navigation, hydropower production, recreation and water quality can be modelled using mass balances in one-dimensional models with time steps of up to a season in length. Flood-flow conditions on the other hand require higher dimension models and much smaller time steps. Groundwater flows can be modelled with much longer time steps than surface water flows but, if both surface and groundwater bodies interact, some way has to be developed to make these different time scales compatible.

Figure 11.1 attempts to identify the range of spatial and temporal scales usually considered by those who model these various components and processes. Much depends, however, on just what impacts are to be identified and on which variables are considered unknown and which are assumed known. For example, it is not uncommon to model water supply systems that include reservoirs and diversions to water consumers on a monthly or even seasonal time-step scale. However, if it is important to capture the within-month changes in reservoir storage for recreation or hydropower, smaller time steps may be necessary during those months or seasons where the variation in storage and flows is important. If water quality impacts are being estimated, time steps of a day or less may be needed, depending on the detail or precision desired. Some watershed processes, such as erosion and evapotranspiration, are best modelled on an hourly or minute-by-minute basis if sufficient data are available to justify those short time steps, whereas other watershed processes, such as changing ownership and land cover, may be modelled using ten-year or longer time steps. Much depends on the variability of both the meteorological and hydrological inputs (air temperatures, radiation, precipitation and streamflows) and the output targets (demands). If there is not much variation during a particular season, it may not be necessary to define smaller time periods within that season.

Figure 11.1 shows the ranges of spatial and temporal scales usually considered when modelling the processes taking place in the three major river basin components – watersheds, surface water bodies and groundwater aquifers. There are exceptions. The boundaries of each zone shown in Figure 11.1 are not as clearly defined as the figure would suggest.

To provide some perspective on spatial extent of some rivers and their watersheds, Table 11.1 lists a few of the

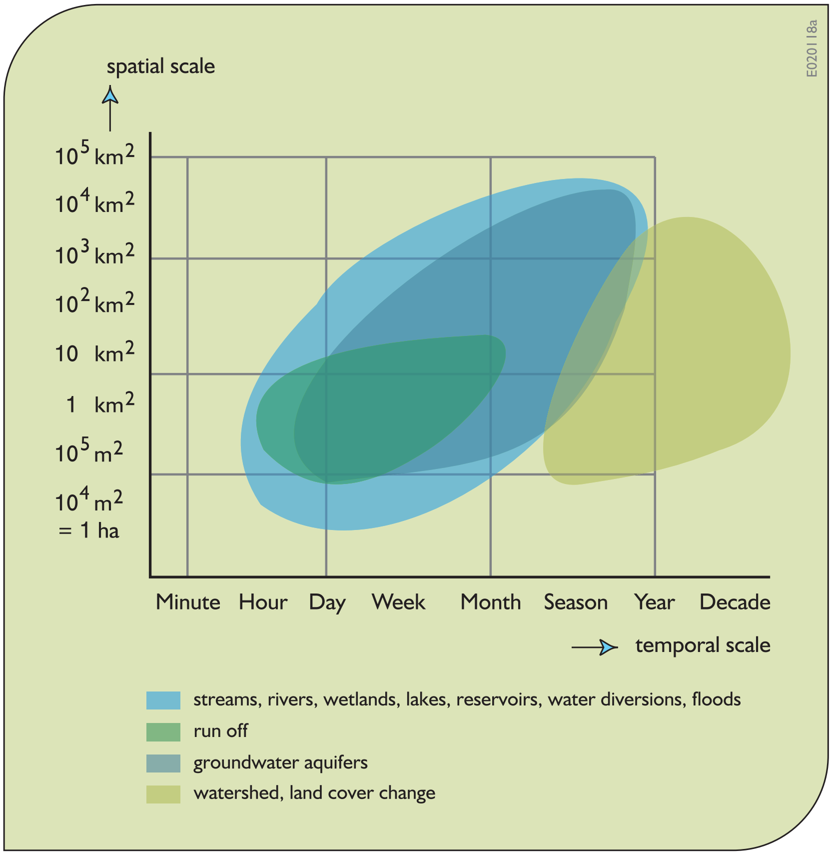



**Figure 11.1.** Common spatial and temporal scales of models of various river basin processes.

world's major rivers with their watershed areas and lengths.

## 1.2. Model Time Periods

When analysing and evaluating various water management plans designed to distribute the natural unregulated flows over time and space, it is usually sufficient to consider average conditions within discrete time periods. Commonly in river basin modelling, weekly, monthly or seasonal flows are used as opposed to daily flows. The shortest time-period duration usually considered in models developed for planning analyses is one that is no less than the time water takes to flow from the upper to the lower end of the river basin. In this case stream and river flows can be defined by simple mass-balance or continuity equations. For shorter duration time periods, some kind of flow routing may be required.

The actual length of each within-year period defined in a model may vary from period to period. The appropriate number and durations of modelled time periods will depend, in part, on the variation of the supply of natural flows and the variation of demands for water flows or storage volumes.

Another important factor to consider in making a decision as to the number and duration of periods to include in any model is the purpose for which the model is to be used. Some analyses are concerned only with identifying the desirable designs and operating policies of various engineering projects for managing water resources at some fixed time (a year, for instance) in the future. Multiple years of hydrological records are simulated to obtain an estimate of just how well a system might perform, at least in a statistical sense, in that future time period. The within-year period durations can have an impact on those performance indicator values, as well as on the estimate of over-year as well as within-year storage requirements that may be needed. These static analyses are not concerned with investment project scheduling or sequencing. Dynamic planning models include changing economic, environmental and other objectives and design and operating parameters over time. As a result, dynamic

E020118b

| river | continent | area $10^3$ km$^2$ | length km |
|---|---|---|---|
| Amazon | S. America | 6915.0 | 6280 |
| Congo | Africa | 3680.0 | 4370 |
| La Plata | S. America | 3100.0 | 4700 |
| Ob Asia | Asia | 2990.0 | 3650 |
| Mississippi | N. America | 2980.0 | 6420 |
| Nile | Africa | 2870.0 | 6670 |
| Yenisei | Asia | 2580.0 | 3490 |
| Lena | Asia | 2490.0 | 4400 |
| Niger | Africa | 2090.0 | 4160 |
| Amur | Asia | 1855.0 | 2820 |
| Yangtze | Asia | 1800.0 | 5520 |
| Mackenzie | N. America | 1790.0 | 5472 |
| Gangas | Asia | 1730.0 | 3000 |
| Volga | Europe | 1380.0 | 3350 |
| Zambezi | Africa | 1330.0 | 2660 |
| St.Lawrence | N. America | 1030.0 | 3060 |
| Orinoco | S. America | 1000.0 | 2740 |
| Indus | Asia | 960.0 | 3180 |
| Yukon | N. America | 850.0 | 3000 |
| Danube | Europe | 817.0 | 2860 |
| Mekong | Asia | 810.0 | 4500 |
| Hwang Ho | Asia | 745.0 | 4670 |
| Columbia | N. America | 668.0 | 1950 |
| Kolyma | Asia | 647.0 | 2130 |
| Sao-Francisco | S. America | 623.0 | 2800 |
| Dnepr | Europe | 504.0 | 2200 |
| Chutsyan | Asia | 437.0 | 2130 |
| Indigirka | Asia | 360.0 | 1726 |
| N.Dvina | Europe | 357.0 | 744 |
| Pechora | Europe | 322.0 | 1810 |
| Godavari | Asia | 314.0 | 1500 |
| Neva | Europe | 281.0 | 74 |
| Magdalena | S. America | 260.0 | 1530 |
| Krishna | Asia | 256.0 | 1290 |
| Fraser | N. America | 233.0 | 1370 |
| Ogowe | Africa | 210.0 | 850 |
| Essequibo | S. America | 155.0 | 970 |
| Sanaga | Africa | 135.0 | 860 |
| Narmada | Asia | 102.0 | 1300 |
| Rhine | Europe | 99.0 | 810 |
| Ebro | Europe | 86.8 | 930 |
| Atrato | S. America | 32.2 | 644 |
| San Juan | S. America | 21.5 | 430 |

**Table 11.1.** Some of the world's major rivers and their spatial characteristics.

models generally span many more years than do static models, but they may have fewer within-year periods.

### 1.3. Modelling Approaches For River Basin Management

Chapter 3 presented a general overview of modelling methods. The distinction made in that chapter between simulation and optimization approaches applies also to river basin planning. The complexity involved in river basin planning requires both approaches. In the design phase of planning studies, optimization methods are often useful for making a first selection among the countless options for capacities and management measures. However, applying optimization requires simplifications, in particular with respect to spatial and temporal detail. For this reason, simulation models should be applied to check and refine the infrastructural designs and operating policies as well as to develop detailed management options. Reference is made to Chapter 3 (Section 3.4) for a more detailed discussion of simulation and optimization models.

Both optimization and simulation modelling approaches require a description of the natural resource system and related infrastructure, and a description of the socio-economic functions in the river basin. These model components will be introduced in the next two sections. In Section 4 these two components are combined for more comprehensive river basin analyses.

## 2. Modelling the Natural Resource System and Related Infrastructure

The natural resource system (NRS) comprises all natural physical, chemical and biological aspects and related artificial elements of the water system. This chapter focuses on the quantitative aspects; water quality and ecology will be described in more detail in Chapter 12. The hydrological surface water and groundwater system will be described, as well as the function of reservoirs to augment the supply in periods of low rainfall and river flows and to contain flood flows in periods of high river flows.

### 2.1. Watershed Hydrological Models

Hydrological modelling is used to predict runoff from land areas, infiltration into soils and percolation into aquifers. Rainfall–runoff models are often used when streamflow gauge data are not available or not reliable, or cannot be made representative of natural flow conditions (that is, what the flows in a stream or river would be without upstream diversions or reservoirs that alter the flows downstream). They are also used to provide estimates of the impact that changing land uses and land covers have on the temporal and spatial distribution of runoff.

### 2.1.1. Classification of Hydrological Models

Hydrological models are classified as either *theoretical* or *empirical* models. A theoretical model is based on physical principles. If all the governing physical processes are described by mathematical functions, a model containing those functions is a physically based model. However, most existing hydrological models simplify the physics and often include empirical components. For example, the conservation of momentum equation or Manning's equation for predicting surface flow include empirical hydraulic resistance terms. Darcy's equation, used to predict subsurface flows, requires an empirical hydraulic conductivity parameter value. Thus they are considered at least partially, if not fully, empirical. Purely empirical or statistical models (Chapter 6) omit the physics and are in reality representations of the observed data.

Depending on the character of the results obtained, hydrological models can also be classified as *deterministic* or *stochastic*. If one or more of the variables in a mathematical model are regarded as random variables whose values can change unpredictably over time, then the model is stochastic. If all the variables are considered to be free from random variation, the model is deterministic. Of course some 'deterministic models' may include stochastic processes that capture some of the spatial and temporal variability of some of the sub-processes, such as infiltration. Most hydrological models are deterministic in spite of a host of random processes taking place in the watersheds to which they are applied. The rather simple modelling approach outlined below for estimating the relative surface and groundwater runoff to surface water bodies is an example of a deterministic model.

Hydrological models can also be classified as *event-based models* or as *continuous-time models*. An event-based model simulates a single runoff event, such as a single storm, usually occurring over a period of time ranging from about an hour to several days. The initial conditions in the watershed for each event must be assumed or determined by other means and supplied as input data. The accuracy of the model output may depend on the reliability of these initial conditions. A continuous-time watershed model includes a sequence of time periods and for each period determines the state of the watershed, whether or not any events take place that will produce surface runoff. The model keeps a continuous account of the watershed surface and groundwater conditions. The effect of any assumed initial conditions decreases rapidly as time advances. Most continuous watershed models include three water balances – one for surface water, another for unsaturated zone moisture content, and a third for groundwater. An event model may omit one or both of the subsurface components and also evapotranspiration when those losses are small in relation to the surface runoff.

Finally, models that have been developed to simulate hydrological processes can be classified as either *lumped* or *distributed* or a mix of both. Lumped models assume homogenous or average conditions over all or portions of a watershed. Lumped models are not sensitive to the actual locations of the varying features in the watershed. Distributed models take into account the locations of various watershed conditions such as land covers, soil types and topography when estimating the total runoff. Both types of models are useful. Distributed models require more detailed data, but are needed if, for example, one wishes to evaluate the impacts of riparian (tree-belt) buffers along streambanks, or the effects of varying topographic features within the watershed. Some models are mixes of the two types of models, in other words, quasi- or semi-distributed models made up of multiple connected lumped models representing different parts of watersheds.

### 2.1.2. Hydrological Processes: Surface Water

The manner and detail in which various hydrological processes are modelled varies among different models and their computer programs. Models that are developed to better understand the physical and other natural processes taking place on and under the surface of a watershed are usually much more detailed than models used for sizing drainage ditches or culverts or for evaluating alternative watershed land use management policies or practices. Here only the basic processes will be outlined. This will identify the minimum types of data needed to model watershed runoff processes.

Figure 11.2 provides a schematic of the hydrological cycle, identifying the processes that are either inputs, such as precipitation and air temperature, and those that are modelled, such as snowmelt or freezing, infiltration, transpiration, percolation, evaporation, and surface and groundwater runoff.

The surface water runoff from watershed land surfaces is affected by many factors. Among them are the amounts and intensity of precipitation and its rain and snow or ice content, the land cover, soil type and slope, and the

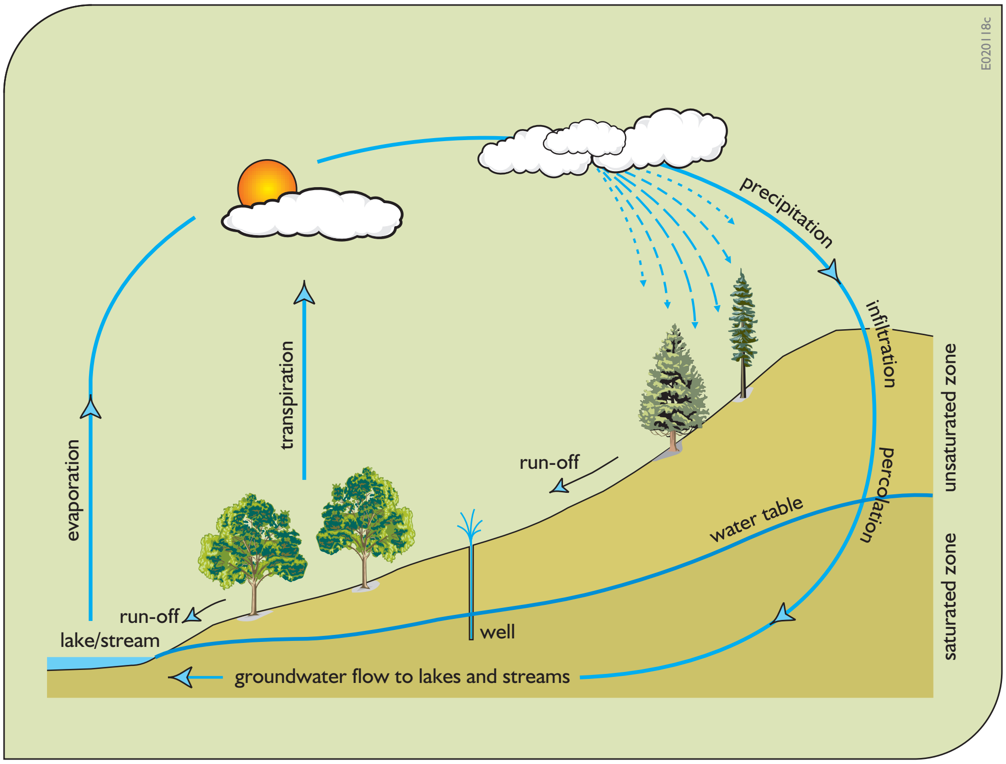



**Figure 11.2.** A schematic diagram of the hydrological cycle applicable to a watershed.

infiltration, evaporation and transpiration that take place. Infiltration, evaporation and transpiration rates are dependent on the land cover, soil type, soil moisture content and air temperature. Both the erosion that takes place and the sediment in the runoff depend on the intensity of precipitation, the land cover, the soil type and the slope. The concentration of chemical constituents in the runoff will depend on the accumulation of chemicals on the land at the time of a storm, the extent they dissolve in the surface water and the extent they are attached to the soil particles that are contained in the runoff. Here, we focus only on the runoff itself, as depicted in Figure 11.2.

The groundwater contribution of runoff to surface water bodies will depend on the relationship between the groundwater and the surface water body. Groundwater runoff can continue long after the last storm may have ended.

Rainfall runoff processes are complex, and depend on the particular physical and biological characteristics of each watershed. One of the difficulties in developing accurate predictive models of rainfall–runoff relationships is the considerable amount of data one must collect to understand just what processes are taking place at rather small spatial and temporal scales and then quantify these processes. Very often such data are just not available, yet we still need some way to predict, at least in some relative sense, the quantities and qualities of surface runoff associated with alternative land use policies or practices and precipitation events.

Ideally, one would like to have continuous records of net precipitation and air temperature for each watershed or sub-watershed being modelled. Usually the hydrological and meteorological data generally available are daily averages. Thus most continuous-time rainfall–runoff models use daily time steps, accepting that this does not capture the intensities of shorter duration storms during the day. A half-hour cloud burst from a thunderstorm over a portion of a modelled watershed may generate much more surface runoff (and sediment loads) than the same amount of precipitation evenly distributed over the entire watershed being modelled over a twenty-four-hour period. The use of daily averages over larger areas will not predict runoff accurately, but they are often the most detailed data one can expect to obtain in practice. More detailed rainfall data can sometimes be obtained from radar images, but the processes of calibrating and analysing such images are relatively expensive.

### Precipitation

Precipitation can be in liquid and/or frozen form, depending of course on the location and time of year. In this discussion the term rain will denote liquid precipitation and snow will denote frozen precipitation (which could include ice as well). The proportion of rain and snow in precipitation will depend on the temperature of the air through which the precipitation falls. The proportion of rain and snow, if not directly measured, usually has to be estimated from average daily surface air temperature data, since these are the most detailed data available. Clearly air temperatures will vary over a day, as well as with altitude. The range of sub-freezing air temperatures over time and at different altitudes can affect the proportion of snow and ice in precipitation as well as the fraction of water content in the snow and ice. Estimates of these proportions, often based on average daily surface temperatures at one or at best only a few measuring stations, are sources of error when direct measurements of snow depth (water equivalent) and rain are not available.

The total net depth of precipitation, $P_t$, in day $t$, equals the sum of the net rain $R_t$ and snow $S_t$ (water equivalent) depths that reach the ground. These are input data for watershed runoff models.

### Surface Water

The water available for surface runoff during any day is the free water already on the ground surface plus the rain and the snowmelt less any losses or reductions due to infiltration and evaporation and any freezing of standing surface water. The depth of snow that melts or the amount of freezing of free-standing water on the ground's surface will depend on the depth of snow available for melting, or on the depth of free surface water available for freezing, and on the air temperature.

Total snow depth accumulation, measured in water content, is determined from a mass balance of snow in the watershed. The initial snow depth, $SD_t$, at the beginning of day $t$ plus the added depth of snow precipitation, $S_t$, and the depth of surface water that freezes on the surface, $WF_t$, less the snowmelt, $SM_t$, equals the final snow depth, $SD_{t+1}$, at the end of each day $t$:

$$SD_{t+1} = SD_t + S_t + WF_t - SM_t \tag{11.1}$$

The actual daily snowmelt, $SM_t$, will be the maximum snowmelt associated with the average air temperature of that day, $SM_t^{\max}$, or the actual amount of snow available for melting, $SD_t + S_t$:

$$SM_t = \min\{SM_t^{\max}, SD_t + S_t\} \tag{11.2}$$

Similarly, for the freezing of surface water, $SW_t$ and rainfall $R_t$,

$$WF_t = \min\{WF_t^{\max}, SW_t + R_t\} \tag{11.3}$$

One would not expect melting and freezing to occur on the same day in a model based on average daily temperatures.

The net amount of water available for runoff, $AW_t$, before other losses are considered, equals the depth available at the beginning of the day, $SW_t$, plus the rainfall depth, $R_t$, and snowmelt depth, $SM_t$, less the freezing depth, $WF_t$:

$$AW_t = SW_t + R_t + SM_t - WF_t \tag{11.4}$$

This amount of water is reduced by that which infiltrates into the unsaturated soil, if any, and by that which evaporates.

### Infiltration and Percolation

Infiltration may occur if the ground is not frozen or completely saturated. The ground is usually considered frozen if snow is on top of it (that is, the accumulated snow depth is greater than zero). It could also be frozen after several days of air temperatures below 0°, even if no snow depth exists. In these cases infiltration is assumed to be zero. If infiltration can occur, the amount will depend on the soil cover, type and moisture content, as well as on the available water to infiltrate if that amount is less than the maximum infiltration rate over a day. Different soil types have different effective porosities (fraction of total soil volume occupied by pores that water can enter), and different maximum rates of infiltration depending, in part, on the soil moisture content.

The maximum soil moisture content of unsaturated soil is reached when its available pore space is occupied by water. This amount, expressed as depth, is equal to the effective porosity of the soil times the depth of the unsaturated zone, $UD_t$. Expressing soil moisture content variable, $MC_t$, as a fraction of the soil's maximum capacity,

Depth of soil moisture at beginning of day $t$,

$$= MC_t(\text{effective porosity})(UD_t) \tag{11.5}$$

The average depth of the unsaturated zone, $UD_t$, approximates the difference between the average watershed surface elevation, $WE$, and the average unconfined groundwater table elevation, $GE_t$:

$$UD_t = WE - GE_t \tag{11.6}$$

The infiltration depth, $I_t$, occurring when the ground is not frozen and free of snow depth, will be the maximum rate, $I^{max}$, for the soil type, the soil moisture deficit, or the water depth available for infiltration, $AW_t$, whichever is less.

$$I_t = \min\{AW_t, I^{max}, [(\text{effective porosity})(UD_t)(1-MC_t)]\}$$
$$\text{if soil not frozen and } SD_t = 0$$
$$= 0 \text{ otherwise} \tag{11.7}$$

The depth of percolation of water, $UP_t$, in the unsaturated soil zone to the groundwater aquifer is assumed to equal either the maximum daily percolation depth, $P^{max}$, for the soil type or the available moisture content down to some minimum percentage, $P^{min}$, of the maximum amount, whichever is less. Often $P^{max}$ is assumed to equal the maximum infiltration rate, $I^{max}$, for lack of any better assumption or measured data.

$$UP_t = \min\{P^{max}, \max[0, (MC_t - P^{min}) \times(\text{effective porosity})(UD_t)]\} \tag{11.8}$$

## Transpiration and Evaporation

Transpiration depends on the moisture content in the unsaturated soil, the land cover, the depth of plant roots and the air temperature. Land cover and root depth may vary over a year.

Plant transpiration, $T_t$, reduces the soil moisture content of the unsaturated zone if the depth of that zone, $UD_t$, exceeds the average root zone depth, $RZ_t$. If the unsaturated zone depth is less than the average root zone depth, then roots have access to the groundwater and the transpiration will reduce the groundwater table. In this case, the transpiration from groundwater, $TG_t$, is usually assumed equal to the maximum transpiration rate, $T_t^{max}$.

Some watershed rainfall–runoff models keep track of the root depths, especially in agricultural croplands, and their impact on soil moisture and transpiration. Alternatively, one can assume transpiration will equal its maximum rate, $T_t^{max}$, for the given land cover and air temperature until the soil moisture content fraction, $MC_t$, goes below some minimal moisture content fraction for transpiration, $MCT^{min}$. Transpiration will cease when soil moisture contents go below that minimum percentage. This water is not available to the plants.

$$T_t = \min\{T_t^{max}, \max[0, (MC_t - MCT^{min}) \times(\text{effective porosity})(UD_t) - UP_t\}]$$

and

$$TG_t \text{ is } 0 \text{ if } UD_t \geq RZ_t \tag{11.9}$$

otherwise

$$TG_t = T_t^{max} \text{ and } T_t \text{ is } 0 \quad \text{if } UD_t < RZ_t \tag{11.10}$$

The depth of evaporation of water on the land surface is a function of the air temperature, relative humidity and wind velocities. Maximum daily evaporation depths are usually assumed to be fixed values that depend on average meteorological conditions for various seasons of the year when more precise data or functions defining evaporation are not available or not used. The actual evaporation each day will be the maximum evaporation rate for that day, $E_t^{max}$, or the total depth of free surface water, $AW_t$, less infiltration, $I_t$, whichever is less.

$$E_t = \min\{AW_t - I_t, E_t^{max}\} \tag{11.11}$$

## Surface Runoff

The depth of surface water available for runoff, $SA_t$, in each day $t$ equals the depth of free surface water, $AW_t$, (which results from the initial amount, $SW_t$, plus rain, $R_t$, and snowmelt, $SM_t$, less the amount that freezes, $WF_t$, as defined in Equation vation. 11.4) less the infiltration $I_t$, and evaporation, $E_t$:

$$SA_t = AW_t - I_t - E_t \tag{11.12}$$

The depth of surface runoff, $SR_t$, is some fraction of this amount, depending on the slope, the extent of ponding, the surface area and the land cover of the watershed.

Surface runoff can be estimated in a number of ways. (See, for example, the curve number method of USDA, 1972, as presented in Chapter 13). For this discussion, assume Manning's Equation applies. This is an equation for determining the velocity of overland and channel flow based on the hydraulic radius, $HR_t$, (cross-sectional area of flow divided by the wetted perimeter), land slope and ground cover. For overland sheet flow, and for channel flow of small depths, the hydraulic radius is essentially the depth of available surface water, $SA_t$. If $SA_t$ is in units of metres, the velocity, $V_t$, of surface runoff (metres/second) based on Manning's equation, equals

$$V_t = (1/n)(SA_t)^{2/3}(\text{Slope})^{1/2} \quad (11.13)$$

The parameter $n$ represents an effective roughness coefficient that includes the effect of raindrop impact, drag over the ground surface from obstacles such as litter, crop ridges and rocks, and erosion and transportation of sediments. This roughness coefficient will be the major calibration parameter of the runoff model.

Watershed runoff will be a combination of sheet flow and channel flow. Typical $n$ values for sheet flow range from 0.011 for smooth paved surfaces to 0.1–0.4 for grass to 0.4–0.8 for underbrush in wooded areas. These $n$ values are for very shallow flow depths of about 20–50 millimetres or so. For river channels the $n$ values typically range between 0.02 and 0.07 (USDA-SCS, 1972).

The maximum surface water flow, $Q_t^{\max}$, equals the velocity times the cross-sectional area of the flow perpendicular to the direction of flow. The cross-sectional area is the depth of available water, $SA_t$, times the average width, $W$, of the watershed. If the units of width are expressed in metres, the maximum flow in cubic metres per second is

$$Q_t^{\max} = (1/n)(SA_t)^{5/3}(\text{Slope})^{1/2}W \quad (11.14)$$

The total quantity of surface water runoff, $SR_t$, from the watershed in day $t$ will be this maximum flow rate multiplied by the number of seconds in a day or the total amount of water available for runoff, whichever is less.

$$SR_t = \min\{Q_t^{\max}(60)(60)(24), (SA_t)(\text{Watershed area})\} \quad (11.15)$$

In reality, most watersheds are not uniform, flat, tilted landscapes where sheet flow takes place. The water available for runoff will either seep into the upper soil layer and reappear down slope or flow into small channels. The above equations based on average watershed conditions are thus a surrogate for what actually takes place at much smaller space scales than are reasonable to model.

### 2.1.3. Hydrological Processes: Groundwater

#### Groundwater Runoff to Streams and Rivers

To estimate the groundwater contribution of the total runoff, it is necessary to model the surface water–groundwater interaction. As shown in Figure 11.3, groundwater can move along flow paths of varying lengths from areas of recharge to areas of discharge. The generalized flow paths in Figure 11.3 start at the water table of the upper unconfined aquifer and continue through the groundwater system, terminating at the surface water body. In the uppermost, unconfined aquifer, flow paths near the

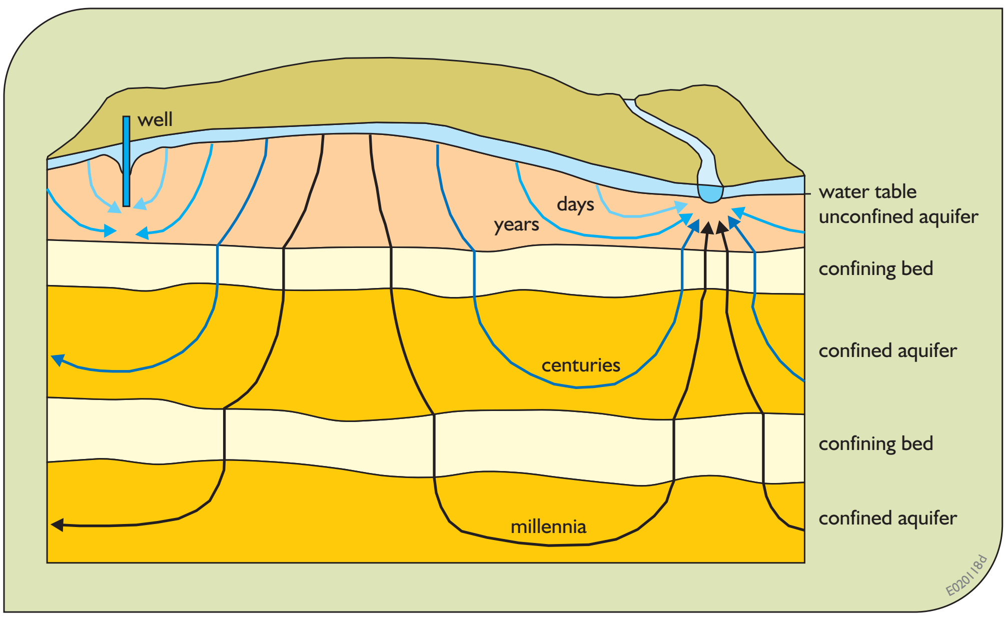



**Figure 11.3.** Groundwater–surface water interactions over a range of time and space scales.

stream can be tens to hundreds of metres in length and have corresponding travel times ranging from days to several years. The longest and deepest flow paths in Figure 11.3 may be thousands of kilometres in length, and travel times may range from decades to millennia. In general, shallow groundwater is more susceptible to contamination from human sources and activities because of its close proximity to the land surface.

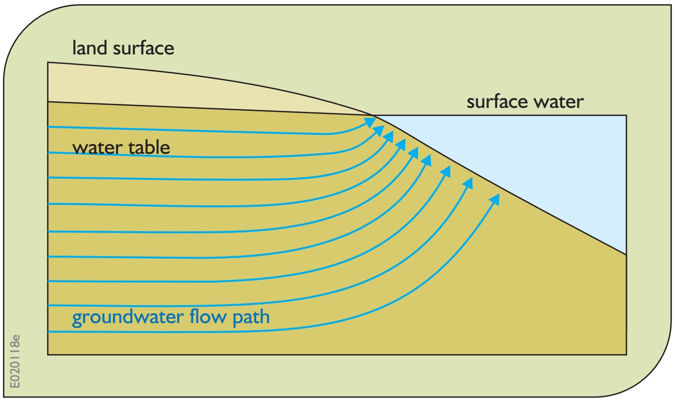



**Figure 11.4.** Groundwater seepage into surface water usually is greatest near shore. In flow diagrams such as that shown here, the quantity of discharge is equal between any two flow lines; therefore, the closer flow lines indicate greater discharge per unit of bottom area.

Streams interact with groundwater in three basic ways. They gain water from inflow of groundwater through the streambed (gaining stream, Figure 11.5A), they lose water to groundwater by outflow through the streambed (losing stream, Figure 11.6A), or they do both, gaining in some reaches and losing in others. For groundwater to flow into a stream channel, the elevation of the groundwater table in the vicinity of the stream must be higher than that of the stream-water surface. Conversely, for surface water to seep to groundwater, the elevation of the water table in the vicinity of the stream must be lower than that of the stream-water surface. Contours of water-table elevation indicate gaining streams by pointing in an upstream direction (Figure 11.5B), and indicate losing streams by pointing in a downstream direction (Figure 11.6B) in the immediate vicinity of the stream.

Losing streams can be connected to the groundwater system by a continuous saturated zone (Figure 11.6A) or can be isolated from the groundwater system by an

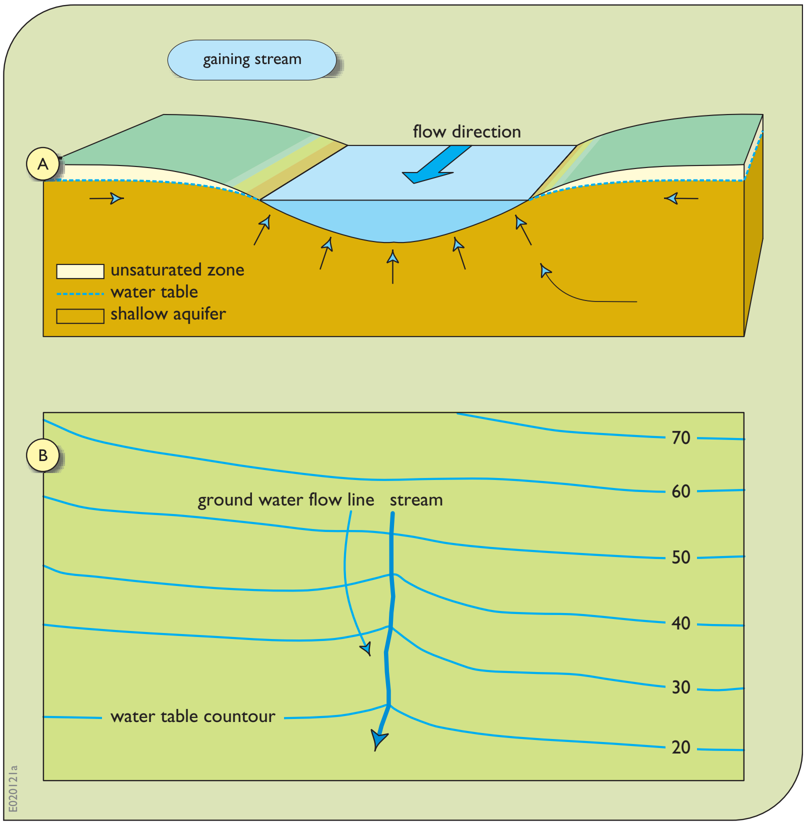



**Figure 11.5.** Gaining streams receive water from the groundwater system (A). This can be determined from water-table contour maps because the contour lines point in the upstream direction where they cross the stream (B).

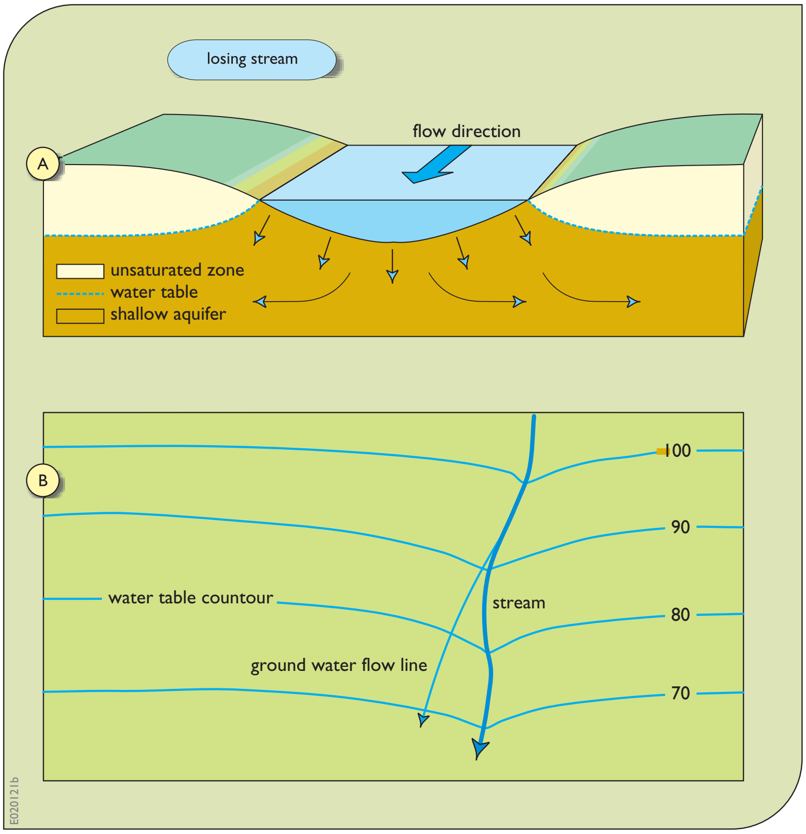



**Figure 11.6.** Losing streams lose water to the groundwater system (A). This can be determined from water-table contour maps because the contour lines point in the downstream direction where they cross the stream (B).

unsaturated zone. In the latter case, infiltration and percolation processes apply. Pumping of shallow groundwaters near disconnected streams does not affect the flow of water in those streams near the pumped wells.

Given the complexity of the interactions between surface water and groundwater at specific sites under different conditions, it becomes difficult to model these interactions without considerable site-specific data based on many detailed observations and measurements. As a first approximation, Darcy's equation for laminar (slowly moving) flow in saturated porous soils can be used to estimate the flow either from the surface water to the groundwater aquifer, or vice versa. This flow depends on the hydraulic conductivity of the soil, $K$ (having dimensions of length per unit time, L/T), and the area through which the flow occurs, $A$ ($L^2$). It also depends on the groundwater flow head gradient, $dH/dX$ (the change in groundwater head $H$ per change in distance $X$ perpendicular to the stream or river reach). Darcy's equation states that the groundwater flow is the product of $KA(dH/dX)$.

Let $(dH/dX)_L$ and $(dH/dX)_R$ be the groundwater head gradients in the direction of groundwater flow on the left and right side of the stream or river reach, respectively. Using Darcy's equation, the net flow, $Q_{sg}$, from surface water stream or river reach to the unconfined groundwater aquifer is the sum of the products of the hydraulic conductivity $K$, the area $A$, and the groundwater head gradient, $dH/dX$, for each side of the stream or river reach.

$$Q_{sg} = K_L A_L (dH/dX)_L - K_R A_R (dH/dX)_R \tag{11.16}$$

If $Q_{sg}$ is negative, the net flow is from the groundwater aquifer to the surface water reach. Clearly, no more water can flow from one water body to another than the volume of water in the former. Models that use equations such as Equation 11.16 need to check on this condition.

## Groundwater Runoff to Lakes, Reservoirs and Wetlands

Surface water in lakes, wetlands and reservoirs interacts with groundwater rather the same way that streams and rivers do, as illustrated in Figure 11.7. Although the basic interactions are the same for these storage water bodies as they are for streams, however, the specific interactions differ in several ways.

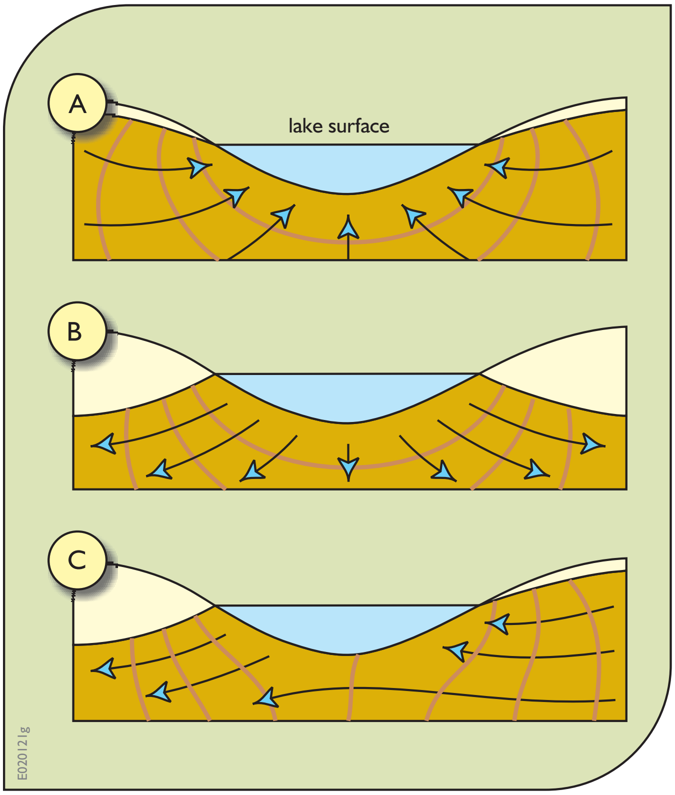



**Figure 11.7.** Lakes can receive groundwater inflow (A), lose water as seepage to groundwater (B), or both (C).

The water levels of natural lakes are not controlled by dams. Hence they generally do not change as rapidly as the water levels of streams, and bank storage is therefore of lesser importance in lakes than in streams. Furthermore, lake sediments commonly have greater volumes of organic deposits. These less permeable organic deposits reduce the seepage and biogeochemical exchanges of water and solutes that take place in lakes compared with those which take place in streams.

Reservoirs are human-made lakes that are designed primarily to control the flow and distribution of surface water. Most reservoirs are constructed in stream valleys, and therefore have some characteristics of both streams and lakes. Like streams, reservoirs can have widely fluctuating levels, bank storage can be significant, and they commonly have a continuous flushing of water through them.

Wetlands that occupy depressions in the land surface have interactions with groundwater similar to lakes and streams. They can receive groundwater inflow, recharge groundwater, or do both. Unlike streams and lakes, however, wetlands do not always occupy low points and depressions in the landscape (Figure 11.8A). They can be present on slopes (fens, for example) or even on drainage divides (some types of bogs). Fens are wetlands that commonly receive most of their water from groundwater discharge (Figure 11.8B). Bogs are wetlands that occupy uplands (Figure 11.8D) or extensive flat areas, and they receive much of their water from precipitation. While riverine wetlands (Figure 11.8C) commonly receive groundwater discharge, they are dependent primarily on the stream for their water supply.

A major difference between lakes and wetlands, with respect to their interaction with groundwater, is the ease with which water moves through their beds. Lakes are commonly shallow around their perimeter, where waves can remove fine-grained sediments, permitting the surface water and groundwater to interact freely. Wetlands, on the other hand, typically contain fine-grained and highly decomposed organic sediments near the wetland edge, slowing the transfer of water and solutes between groundwater and surface water.

Another difference in the interaction between groundwater and surface water in wetlands compared with lakes is the extent of rooted vegetation in wetlands. The fibrous root mat in wetland soils is highly conductive to water flow; therefore, water uptake by roots of emergent plants results in significant interchange between surface water and pore water of wetland sediments. Water exchanges in this upper soil zone are usually not as restricted as the exchanges between surface water and groundwater at the base of the wetland sediments.

All these complex interactions between ground and surface water systems can only be approximated by watershed models, even the most detailed ones. Without extensive data, it is unlikely the models will capture all the interactions that take place over space and time.

### 2.1.4. Modelling Groundwater: Surface Water Interactions

One of the simplest ways of estimating the groundwater runoff from a watershed is to assume the runoff will equal some fraction of the average groundwater head. This fraction is called a *recession constant*.

If the aquifer lies solely under the watershed of interest, then a recession rate can be defined for estimating

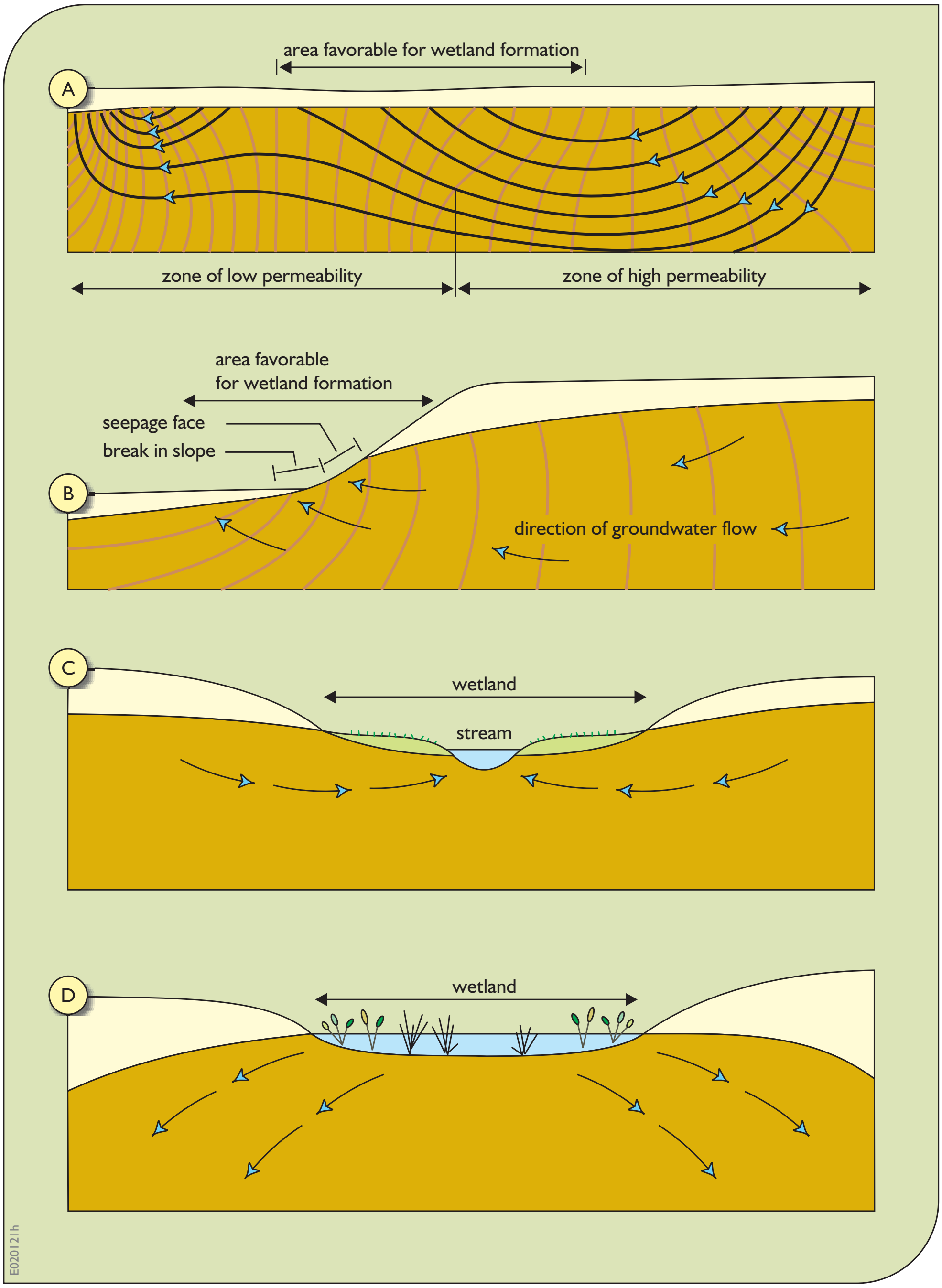



**Figure 11.8.** The source of water to wetlands can be from groundwater discharge where the (light-coloured) land surface is underlain by complex groundwater flow fields (A), from groundwater discharge at seepage faces and at breaks in slope of the water table (B), from streams (C), and from precipitation in cases where wetlands have no stream inflow and groundwater gradients slope away from the wetland (D).

the extent of groundwater base flow depth when the average groundwater table elevation, $GE_t$, is greater than the average surface water body elevation, $SE_t$. Letting the dimensionless recession fraction be $k$, the groundwater base flow contribution, $GR_t$, in day $t$, to total surface runoff is

$$GR_t = \text{Max}\{0, (GE_t - SE_t\}k(\text{aquifer area}) \times(\text{effective aquifer porosity}) \qquad (11.17)$$

The total runoff in day $t$ equals the sum of the groundwater runoff, $GR_t$, plus the surface water runoff $SR_t$.

This method does not consider situations in which water flows from the surface water bodies to the groundwater aquifer.

Alternatively, the two-way interactions between the ground and surface waters, perhaps involving multiple watersheds and one or more underlying unconfined aquifers, can be modelled using Darcy's equation (Equation 11.17).

If the groundwater aquifer is not being modelled in detail, all one may have available from a lumped watershed model is the average groundwater table elevation or head at the beginning or end of any time period. This average head is usually based on the groundwater volume. For an unconfined aquifer, the average head, $H_t$, at the beginning of day $t$, will equal the groundwater volume, $GWV_t$, divided by the product of the aquifer area times its effective porosity. The head gradient is the difference between the average groundwater elevation (head), $GE_t$, and the surface elevation, $SE_t$, of the surface water body, divided by some representative distance $\Delta X$.

$$- dH_t/dX \approx [GE_t - SE_t]/\Delta X \qquad (11.18)$$

Many lumped models of groundwater aquifers approximate the head gradient by dividing the elevation difference between the average groundwater table elevation, $GE_t$, and the surface water body elevation, $SE_t$, by no more than a quarter of the distance from the surface water body and the aquifer boundary. This is shown in Figure 11.9 for water-gaining surface water bodies. In this figure, the $\Delta W$ term is equivalent to the $\Delta X$ terms use in Equations 11.18 and 11.19.

The approximate area, $A_t$, through which the groundwater flows is the average of the surface water body elevation above the base of the aquifer, $SE_t$, and the average groundwater head elevation, $GE_t$, times the length, $L$, of the surface water body in its direction of flow. Assuming groundwater flows from both sides of the river into the river and the $K$ and $\Delta X$ (or $\Delta W$ in Figure 11.9) terms are the same for both sides of the river, Darcy's equation for

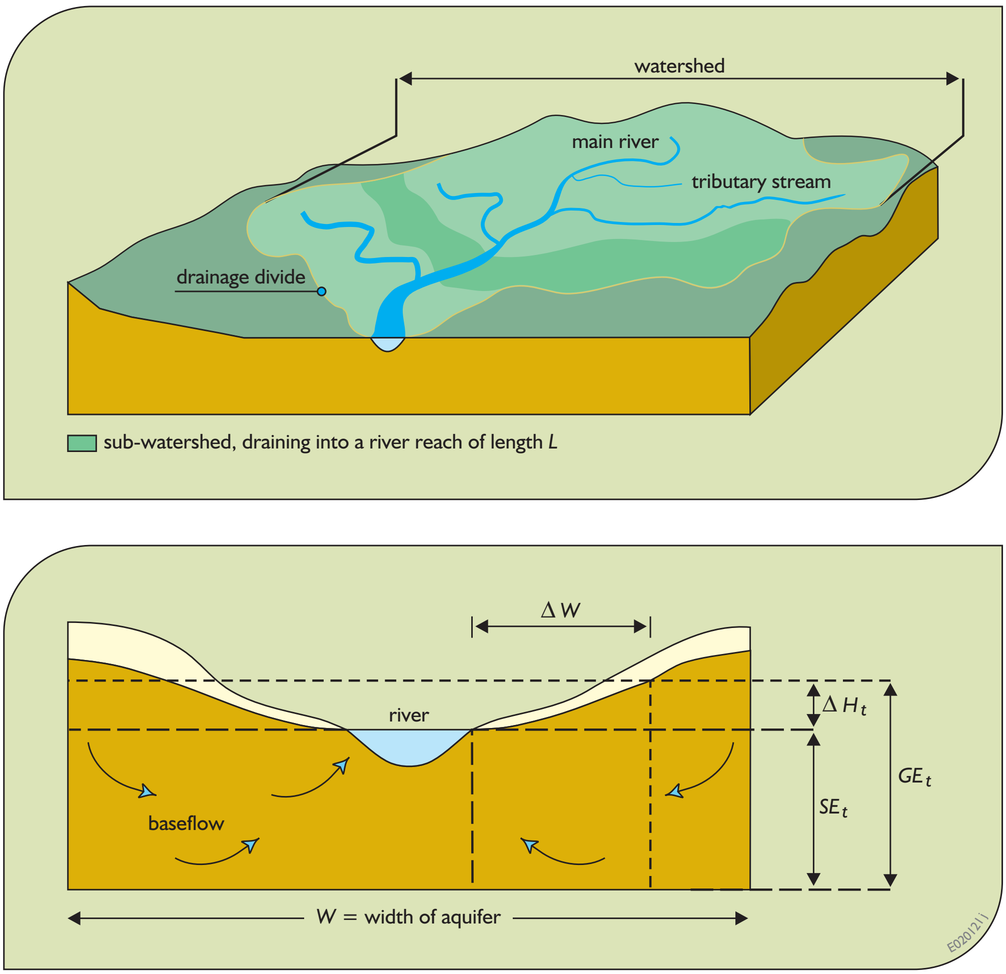



**Figure 11.9.** Cross-sectional dimensions of a watershed showing a surface water body that is receiving water from the groundwater aquifer.

estimating the groundwater runoff, $GR_t$ from this portion of the watershed is:

$$\begin{aligned} GR_t &= 2K((SE_t + GE_t)/2)(L)[GE_t - SE_t]/\Delta X \\ &= K(GE_t^2 - SE_t^2)(L)/\Delta X \end{aligned} \tag{11.19}$$

This equation can apply to any surface water body receiving runoff from groundwater. It can also apply to any groundwater aquifer receiving water from a surface water body. In this case, $GR_t$ would be negative. As for Equation 11.16, no more water can flow from one water body to another than the volume of water in the former. Again, models that use equations such as Equation 11.19 need to check on this condition.

### 2.1.5. Streamflow Estimation

Water resources managers need to have estimates of streamflows at each site where management decisions are to be made. These streamflows can be based on the results of rainfall runoff models or on measured historical flows at stream gauges. For modelling alternative management policies, these streamflows should be those that would have occurred under natural conditions. These are called *naturalized flows*. These naturalized flows are often computed from observed gauge flows adjusted to take into account any upstream regulation and diversions. Many gauge flow measurements reflect actions taken upstream that alter the natural flows, such as diversions and reservoir storage. Unless such upstream water management and use policies are to continue, these regulated flows should be converted to unregulated or natural flows prior to their use in management models.

Alternatively, naturalized flows can be estimated from rainfall–runoff modelling.

#### Rainfall–Runoff Modelling

In rainfall–runoff models, mass balances must be maintained for snow depth, for free surface water depth, for the moisture content of the unsaturated zone, and for the groundwater volume, as applicable.

For surface water, the remaining free surface water will be that water which is left at the end of period $t$. The free water remaining on the watershed's surface at the end of day $t$, $SW_{t+1}$, is what exists at the beginning of day $t+1$. It equals the surface water that exists at the beginning of the day, $SW_t$, plus the rainfall, $R_t$, plus the snowmelt, $SM_t$, less the depth that freezes, $WF_t$, the infiltration, $I_t$, the evaporation, $E_t$, and the depth of surface runoff, $SR_t$/(watershed area):

$$\begin{aligned} SW_{t+1} &= SW_t + R_t + SM_t - WF_t - I_t - E_t \\ &\quad - SR_t/(\text{watershed area}) \\ &= SA_t - SR_t/(\text{watershed area}) \end{aligned} \tag{11.20}$$

The snow depth, $SD_{t+1}$, at the end of each day $t$ is:

$$SD_{t+1} = SD_t + S_t + WF_t - SM_t \tag{11.21}$$

The final groundwater volume is equal to the initial groundwater volume, $GWV_t$, plus any additions and less any reductions during the day. These additions and reductions include the total net runoff to surface water bodies, $GR_t$, the percolation, $UP_t$, the groundwater transpiration, $TG_t$, any groundwater abstractions or artificial recharge, less any losses to, $GWQ_t^{\text{out}}$, or plus any gains from, $GWQ_t^{\text{in}}$, deeper aquifers.

$$GWV_{t+1} = GWV_t + \left[\sum_{\text{watersheds}} GR_t + (UP_t - TG_t) \times (\text{watershed area})\right] + GWQ_t^{\text{in}} - GWQ_t^{\text{out}} \tag{11.22}$$

The average groundwater table elevation is computed taking into account the percolation from the unsaturated zone and its inflow from or outflow to the surface water system. The groundwater head or elevation is the watershed elevation or its volume divided by its area times its porosity, whichever is less:

$$\begin{aligned} GE_{t+1} &= \min\{WE, \{GWV_{t+1}/[(\text{aquifer area}) \\ &\qquad \times(\text{effective porosity})]\}\} \end{aligned} \tag{11.23}$$

The unsaturated zone depth at the end of each day $t$ is

$$UD_{t+1} = WE - GE_{t+1} \tag{11.24}$$

The soil moisture fraction, $MC_{t+1}$, in the unsaturated soil layer at the end of each day can be calculated as

$$\begin{aligned} MC_{t+1} &= [MC_t(\text{porosity})(UD_t) + I_t - T_t - UP_t]/ \\ &\qquad (\text{effective porosity})(UD_{t+1}) \end{aligned} \tag{11.25}$$

If the groundwater table is rising, the actual soil moisture content will probably increase due to capillary action. If the groundwater table is falling, the actual soil moisture content is likely to be higher than estimated due to the delayed percolation of the excess water from the once-saturated zone. Here, we assume water table elevation changes (that is, unsaturated zone depth changes) do not by themselves change the percentage of soil moisture in the unsaturated zone.

If the groundwater table equals the surface elevation, then the unsaturated zone, *UD*, does not exist and no infiltration and percolation will occur. In this case, the unconfined saturated groundwater table is at the surface.

### Streamflow Estimation Based on Flow Data

Assuming unregulated streamflow data are available at gauge sites, these sites may not be where flow data are needed. Thus, some way is needed to generate corresponding streamflow data at sites where managers need them, such as at potential diversion or reservoir sites. Just how this can be done depends in part on the durations of a model's time periods.

Consider, for example, the simple river basin illustrated in Figure 11.10. The streamflows have been recorded over a number of years at gauge sites 1 and 9. Knowledge of the flows $Q_t^s$ in each period $t$ at gauge sites $s = 1$ and 9 permits the estimation of flows at any other site in the basin as well as the incremental flows between those sites in each period $t$.

The method used to estimate flows at ungauged sites will depend on the characteristics of the watershed of the river basin. In humid regions where streamflows increase in the downstream direction, and the spatial distribution of average monthly or seasonal rainfall is more or less the same from one part of the river basin to another, the

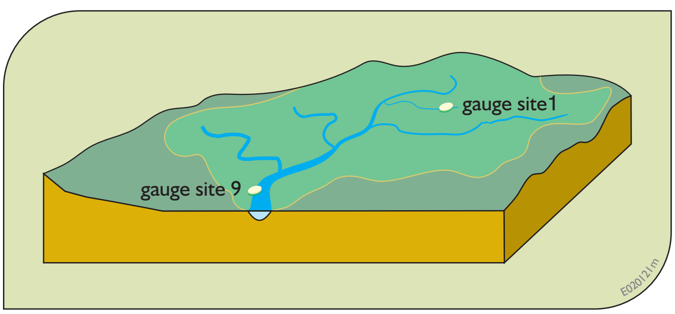



**Figure 11.10.** River basin gauge sites where streamflows are measured and recorded.

runoff per unit land area is assumed constant over space. In these situations, estimated flows, $q_t^s$, at any site $s$ are usually based on the watershed areas, $A^s$, contributing flow to those sites, and the corresponding streamflows and watershed areas above the nearest or most representative gauge sites.

For each gauge site, the runoff per unit land area can be calculated by dividing the gauge flow, $Q_t^s$, by the upstream drainage area, $A^s$. This can be done for each gauge site in the basin. Thus for any gauge site $g$, the runoff per unit drainage area in month or season $t$ is $Q_t^g$ divided by $A^g$. This runoff per unit land area times the drainage area upstream of any site $s$ of interest will be the estimated streamflow in that period at that site $s$. If there are multiple gauge sites, as illustrated in Figure 11.10, the estimated streamflow at some ungauged site $s$ can be a weighted linear combination of those unit area runoffs times the area contributing to the flow at site $s$. The non-negative weights, $w_g$, that sum to 1 reflect the relative significance of each gauge site with respect to site $s$. Their values will be based on the judgements of those who are familiar with the basin's hydrology.

$$Q_t^s = \left\{ \sum_g w_g Q_t^g / A^g \right\} A^s \tag{11.26}$$

In all the models developed and discussed below, the variable $Q_t^s$ will refer to the mean natural (unregulated) flow ($L^3/T$) at a site $s$ in a period $t$.

The difference between the natural streamflows at any two sites is called the *incremental flow*. Using Equation 11.26 to estimate streamflows will result in positive incremental flows. The downstream flow will be greater than the upstream flow. In arid regions, incremental flows may not exist and hence, due to losses, the flows may be decreasing in the downstream direction. In these cases, there is a net loss in flow in the downstream direction. This might be the case when a stream originates in a wet area and flows into a region that receives little if any rainfall. In such arid areas the runoff into the stream, if any, is less than the evapotranspiration and infiltration into the ground along the stream channel.

For stream channels where there exist relatively uniform conditions affecting water loss and where there are no known sites where the stream abruptly enters or exits the ground (as can occur in karst conditions), the average streamflow for a particular period $t$ at site $s$ can be based on the nearest or most representative gauge flow, $Q_t^g$, and a loss rate per unit length of the stream or river, $L^{gs}$, between gauge site $g$ and an ungauged site $s$. If there are at least two gauge sites along the portion of the stream or river that is in the dry region, one can compute the loss of flow per unit stream length, and apply this loss rate to various sites along the stream or river. This loss rate per unit length, however, may not be constant over the entire length between the gauge stations, or even for all flow rates. Losses will probably increase with increasing flows simply because more water surface is exposed to evaporation and seepage. In these cases one can define a loss rate per unit length of stream or river as a function of the magnitude of flow.

In watersheds characterized by significant elevation changes, and consequently varying rainfall and runoff distributions, other methods may be required for estimating average streamflows at ungauged sites. The selection of the most appropriate method to use, as well as the most appropriate gauge sites to use, to estimate the streamflow, $Q_t^s$, at a particular site $s$ can be a matter of judgement. The best gauge site need not necessarily be the nearest gauge to site $s$, but rather the site most similar with respect to important hydrological variables.

### 2.1.6. Streamflow Routing

If the duration of a within-year period is less than the time of flow throughout the stream or river system being modelled, and the flows vary within the system, then some type of streamflow routing must be used to keep track of where the varying amounts of water are in each time period. There are many proposed routing methods (as described in any hydrology text or handbook, e.g. Maidment, 1993). Many of these more traditional methods can be approximated with sufficient accuracy by relatively simple methods. Two such methods are described in the following paragraphs.

The outflow, $O_t$, from a reach of stream or river during a time period $t$ is a function of the amount of water in that reach, i.e. its initial storage, $S_t$, and its inflow, $I_t$. Because of bank storage, that outflow is often dependent on whether the quantity of water in the reach is increasing or decreasing. If bank inflows and outflows are explicitly modelled, for example as described in Sections 2.1.3 and 2.1.4, or if bank storage is not that significant, then the outflow from a reach in any period $t$ can be expressed as a simple two-parameter power function of the form $a(S_t + I_t)^b$. Mass-balance equations, which may take losses into account, update the initial storage volumes in each succeeding time period. The reach-dependent parameters $a$ and $b$ can be determined through calibration procedures such as genetic algorithms (Chapter 6), given a time series of reach inflows and outflows. The resulting outflow function is typically concave (the parameter $b$ will be less than 1), and thus the minimum value of $S_t + I_t$ must be at least 1. If, due to evaporation or other losses, the reach volume drops below this or any pre-selected higher amount, then the outflow is assumed to be 0.

Alternatively, one can adopt a three- or four-parameter routing approach that fits a wider range of conditions. Each stream or river reach can be divided into a number of segments. That number $n$ is one of the parameters to be determined. Each segment $s$ can be modelled as a storage unit, having an initial storage volume, $S_{st}$, and an inflow, $I_{st}$. The three-parameter approach assumes the outflow, $O_{st}$, is a linear function of the initial storage volume and inflow:

$$O_{st} = \alpha S_{st} + \beta I_{st} \tag{11.27a}$$

Equation 11.27a applies for all time periods $t$ and for all reach segments $s$ in a particular reach. Different reaches may have different values of the parameters $n$, $\alpha$ and $\beta$. The calibrated values of $\alpha$ and $\beta$ are non-negative and no greater than 1. Again, a mass-balance equation updates each segment's initial storage volume in the following time period. The outflow from each reach segment is the inflow into the succeeding reach segment.

The four-parameter approach assumes that the outflow, $O_{st}$, is a non-linear function of the initial storage volume and inflow:

$$O_{st}, = (\alpha S_{st} + \beta I_{st})^{\gamma} \tag{11.27b}$$

where the forth parameter $\gamma$ is greater than 0 and no greater than 1. In practice, $\gamma$ is very close to 1. Again the values of these parameters can be found using non-linear optimization methods, such as genetic algorithms, together with a time series of observed reach inflows and outflows.

Note the flexibility available when using the three- or four-parameter routing approach. Even blocks of flow can be routed a specified distance downstream over a specified time, regardless of the actual flow. This can be done

**Box 11.1. Streamflow Routing in RIBASIM Simulation Model**

Streamflow routing in RIBASIM is simulated with link storage nodes. The outflow from these 'nodes' (actually river segments acting as reservoirs) into the next node can be described by:

- Manning formula
- two-layered, multi-segmented Muskingum formula
- Puls method
- Laurenson non-linear 'lag and route' method.

by setting $\alpha$ and $\gamma$ to 1, $\beta$ to 0, and the number of segments n to the number of time periods it takes to travel that distance. This may not be very realistic, but there are some river basin reaches where managers believe this particular routing applies.

Most river basin models offer a number of alternative approaches to describe streamflow routing. Box 11.1 illustrates the methods included in the RIBASIM package (WL | Delft Hydraulics, 2004). Which method is the best to apply will depend on the required accuracy and the available data.

## 2.2. Lakes and Reservoirs

Lakes and reservoirs are sites in a basin where surface water storage needs to be modelled. Thus, variables defining the water volumes at those sites must be defined. Let $S_t^s$ be the initial storage volume of a lake or reservoir at site $s$ in period $t$. Omitting the site index $s$ for the moment, the final storage volume in period $t$, $S_{t+1}$, (which is the same as the initial storage in the following period $t+1$) will equal the initial volume $S_t$ plus the net surface and groundwater inflows, $Q_t$, less the release or discharge, $R_t$, and evaporation and seepage losses, $L_t$. All models of lakes and reservoirs include this mass-balance equation for each period $t$ being modelled:

$$S_t + Q_t - R_t - L_t = S_{t+1} \tag{11.28a}$$

The release from a natural lake is a function of its surrounding topography and its water surface elevation. It is determined by nature, and unless it is made into a reservoir, its discharge or release is not controlled or managed. The release from a reservoir is controllable, and is usually a function of the reservoir storage volume and time of year. Reservoirs also have fixed storage capacities, $K$. In each period $t$, reservoir storage volumes, $S_t$, cannot exceed their storage capacities, $K$.

$$S_t \le K \quad \text{for each period } t. \tag{11.28b}$$

Equations 11.28a and 11.28b are the two fundamental equations required when modelling water supply reservoirs. They apply for each period $t$.

The primary purpose of all reservoirs is to provide a means of regulating downstream surface water flows over time and space. Other purposes may include storage volume management for recreation and flood control, and storage and release management for hydropower production. Reservoirs are built to alter the natural spatial and temporal distribution of the streamflows. The capacity of a reservoir and its release (or operating) policy determine the extent to which surface water flows can be stored for later release.

The use of reservoirs for temporarily storing streamflows often results in a net loss of total streamflow due to increased evaporation and seepage. Reservoirs also bring with them changes in the ecology of a watershed and river system. They may also displace humans and human settlements. When considering new reservoirs, any benefits derived from regulation of water supplies, from floodwater storage, from hydroelectric power, and from any navigational and recreational activities should be compared to any ecological and social losses and costs. The benefits of reservoirs can be substantial, but so may the costs. Such comparisons of benefits and costs are always challenging because of the difficulty of expressing all such benefits and costs in a common metric. For a full discussion on the complex issues and choices on reservoir construction, refer to report of the World Commission on Dams (WCD, 2000).

Reservoir storage capacity can be divided among three major uses:

- *active storage* used for downstream flow regulation and for water supply, recreational development or hydropower production
- *dead storage* required for sediment collection
- *flood storage* capacity reserved to reduce potential downstream flood damage during flood events.

These separate storage capacities are illustrated in Figure 11.11. The distribution of active and flood control storage capacities may change over the year. For example

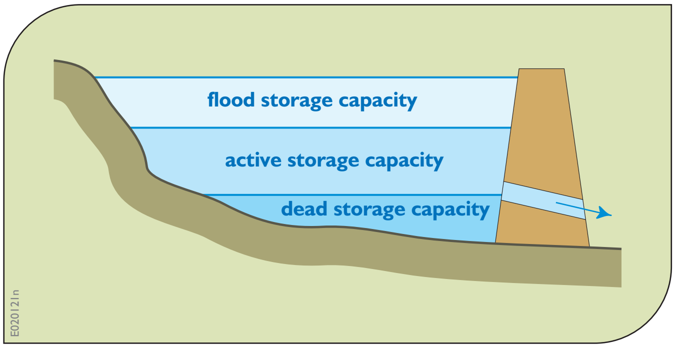



**Figure 11.11.** Total reservoir storage volume capacity consisting of the sum of dead storage, active storage and flood control storage capacities.

there is no need for flood control storage in seasons that are not going to experience floods.

Often these components of reservoir storage capacity can be modelled separately and then added together to determine total reservoir storage capacity. The next several sections of this chapter address how these capacities may be determined.

### 2.2.1. Estimating Active Storage Capacity

#### Mass Diagram Analyses

Perhaps one of the earliest methods used to calculate the active storage capacity required to meet a specified reservoir release, $R_t$, in a sequence of periods $t$, was developed by W. Rippl (1883). His *mass diagram analysis* is still used today by many planners. It involves finding the maximum positive cumulative difference between a sequence of pre-specified (desired) reservoir releases $R_t$ and known inflows $Q_t$. One can visualize this as starting with a full reservoir, and going through a sequence of simulations in which the inflows and releases are added and subtracted from that initial storage volume value. Doing this over two cycles of the record of inflows will identify the maximum deficit volume associated with those inflows and releases. This is the required reservoir storage. Having this initial storage volume, the reservoir would always have enough water to meet the desired releases. However, this only works if the sum of all the desired releases does not exceed the sum of all the inflows over the same sequence of time periods. Reservoirs cannot make water. Indeed, they can lose it from evaporation and seepage.

Equation 11.29 represents this process. The active storage capacity, $K_a$, will equal the maximum accumulated storage deficit one can find over some interval of time within two successive record periods, $T$.

$$K_a = \text{maximum}\left[\sum_{t=i}^{j}(R_t - Q_t)\right] \tag{11.29}$$

where $1 \leq i \leq j \leq 2T$.

Equation 11.29 is the analytical equivalent of graphical procedures proposed by Rippl for finding the active storage requirements. Two of these graphical procedures are illustrated in Figures 11.12 and 11.13 for a nine-period inflow record of 1, 3, 3, 5, 8, 6, 7, 2 and 1. Rippl's original approach, shown in Figure 11.12, involves plotting the cumulative inflow $\sum_{\tau=1}^{t} Q_\tau$ versus time $t$. Assuming a constant reservoir release, $R_t$, in each period $t$, a line with slope $R_t$ is placed so that it is tangent to the cumulative inflow curve. To the right of these points of tangency the release $R_t$ exceeds the inflow $Q_t$. The maximum vertical distance between the cumulative inflow curve and the release line of slope $R_t$ equals the maximum water deficit, and hence the required active storage capacity. Clearly, if the average of the releases $R_t$ is greater than the mean inflow, a reservoir will not be able to meet the demand no matter what its active storage capacity.

An alternative way to identify the required reservoir storage capacity is to plot the cumulative non-negative deviations, $\sum_{\tau}^{t}(R_\tau - Q_\tau)$, and note the largest total deviation, as shown in Figure 11.13.

These graphical approaches do not account for losses. Furthermore, the method shown in Figure 11.12 is awkward if the desired releases in each period $t$ are not the same. The equivalent method shown in Figure 11.13 is called the sequent peak method. If the sum of the desired releases does not exceed the sum of the inflows, calculations over at most two successive hydrological records of flows may be needed to identify the largest cumulative deficit inflow. After that the procedure will produce repetitive results. It is much easier to consider changing as well as constant release values when determining the maximum deficit by using the sequent peak method.

#### Sequent Peak Analyses

The sequent peak procedure is illustrated in Table 11.2. Let $K_t$ be the maximum total storage requirement needed

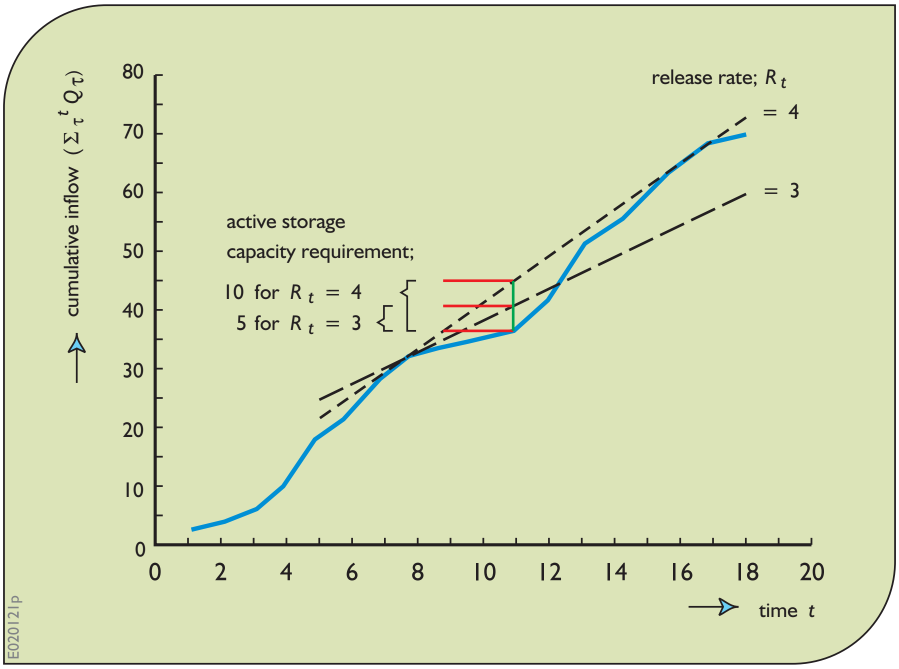



**Figure 11.12.** The Rippl or mass diagram method for identifying reservoir active storage capacity requirements.

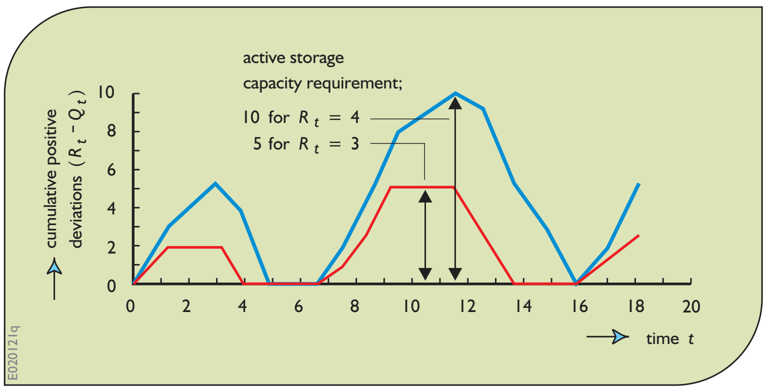



**Figure 11.13.** Alternative plot for identifying reservoir active storage capacity requirements.

for periods 1 up through period $t$. As before, let $R_t$ be the required release in period $t$, and $Q_t$ be the inflow in that period. Setting $K_0$ equal to 0, the procedure involves calculating $K_t$ using Equation 11.30 (below) consecutively for up to twice the total length of record. This assumes that the record repeats itself to take care of the case when the critical sequence of flows occurs at the end of the streamflow record, as indeed it does in the example nine-period record of 1, 3, 3, 5, 8, 6, 7, 2 and 1.

$$\begin{aligned} K_t &= R_t - Q_t + K_{t-1} \quad \text{if positive,} \\ &= 0 \text{ otherwise} \end{aligned} \tag{11.30}$$

The maximum of all $K_t$ is the required storage capacity for the specified releases $R_t$ and inflows, $Q_t$. Table 11.2 illustrates this sequent peak procedure for computing the active capacity $K_a$, i.e. the maximum of all $K_t$, required to achieve a release $R_t = 3.5$ in each period given the series of nine streamflows. Note that this method does not require all releases to be the same.

### 2.2.2. Reservoir Storage–Yield Functions

Reservoir storage–yield functions define the minimum active storage capacity required to ensure a given constant release rate for a specified sequence of reservoir inflows. Mass diagrams, sequent peak analyses and linear optimization (Chapter 4) are three methods that can be used to define these functions. Given the same sequence of known inflows and specified releases, each method will provide identical results. Using optimization models,

| time $t$ | $(R_t - Q_t + K_{t-1})^+ = K_t$ | |
|---|---|---|
| 1 | 3.5 − 1.0 + 0.0 = 2.5 | |
| 2 | 3.5 − 3.0 + 2.5 = 3.0 | |
| 3 | 3.5 − 3.0 + 3.0 = 3.5 | |
| 4 | 3.5 − 5.0 + 3.5 = 2.0 | |
| 5 | 3.5 − 8.0 + 2.0 = 0.0 | |
| 6 | 3.5 − 6.0 + 0.0 = 0.0 | |
| 7 | 3.5 − 7.0 + 0.0 = 0.0 | |
| 8 | 3.5 − 2.0 + 0.0 = 1.5 | |
| 9 | 3.5 − 1.0 + 1.5 = 4.0 | |
| 1 | 3.5 − 1.0 + 4.0 = 6.5 | |
| 2 | 3.5 − 3.0 + 6.5 = 7.0 | |
| 3 | 3.5 − 3.0 + 7.0 = 7.5 | $K_a$ |
| 4 | 3.5 − 5.0 + 7.5 = 6.0 | |
| 5 | 3.5 − 8.0 + 6.0 = 1.5 | |
| 6 | 3.5 − 6.0 + 1.5 = 0.0 | repetition begins |
| 7 | 3.5 − 7.0 + 0.0 = 0.0 | |
| 8 | 3.5 − 2.0 + 0.0 = 1.5 | |
| 9 | 3.5 − 1.0 + 1.5 = 4.0 | |

E020121r

**Table 11.2. Illustration of the sequent peak procedure for computing active storage requirements.**

it is possible to obtain such functions from multiple reservoirs and to account for losses based on storage volume surface areas, as will be discussed later.

There are two ways of defining a linear optimization (linear programming) model to estimate the active storage capacity requirements. One approach is to minimize the active storage capacity, $K_a$, subject to minimum required constant releases, $Y$, the yield. This minimum active storage capacity is the maximum storage volume, $S_t$, required, given the sequence of known inflows $Q_t$, and the specified yield, $Y$, in each period $t$. The problem is to find the storage volumes, $S_t$, and releases, $R_t$, that

$$\text{Minimize } K_a \tag{11.31}$$

subject to:
mass-balance constraints:

$$S_t + Q_t - R_t = S_{t+1} \quad t = 1, 2, \ldots, T; T+1 = 1 \tag{11.32}$$

capacity constraints:

$$S_t \leq K_a \quad t = 1, 2, \ldots, T \tag{11.33}$$

minimum release constraints:

$$R_t \geq Y \quad t = 1, 2, \ldots, T \tag{11.34}$$

for various values of the yield, $Y$.

Alternatively, one can maximize the constant release yield, $Y$,

$$\text{Maximize } Y \tag{11.35}$$

for various values of active storage capacity, $K_a$, subject to the same constraint Equations 11.32 through 11.34.

Constraints 11.32 and 11.34 can be combined to reduce the model size by $T$ constraints.

$$S_t + Q_t - Y \geq S_{t+1} \quad t = 1, 2, \ldots, T; T+1 = 1 \tag{11.36}$$

The solutions of these two linear programming models, using the nine-period flow sequence referred to above and solved for various values of yield or capacity, respectively, are plotted in Figure 11.14. The results are the same as could be found using the mass diagram or sequent peak methods.

There is a probability that the storage–yield function just defined will fail. A record of only nine flows, for example, is not very long and, hence, will not give one much confidence that they will define the critical low-flow period of the future. One rough way to estimate the reliability of a storage–yield function is to rearrange and rank the inflows in order of their magnitudes. If there are $n$ ranked inflows, there will be $n+1$ intervals separating them. Assuming there is an equal probability that any future flow could occur in any interval between these ranked flows, there is a probability of $1/(n+1)$ that a future flow will be less than the lowest recorded flow. If that record low flow occurs during a critical low flow period, more storage may be required than indicated in the function.

Hence, for a record of only nine flows that are considered representative of the future, one can be only about 90% confident that the resulting storage–yield function will apply in the future. One can be only 90% sure of the predicted yield $Y$ associated with any storage volume $K$. A much more confident estimate of the reliability of any derived storage–yield function can be obtained by using synthetic flows to supplement any measured streamflow record, taking parameter uncertainty into account (as discussed in Chapters 7, 8, and 9). This will provide alternative sequences as well as more intervals between ranked flows.

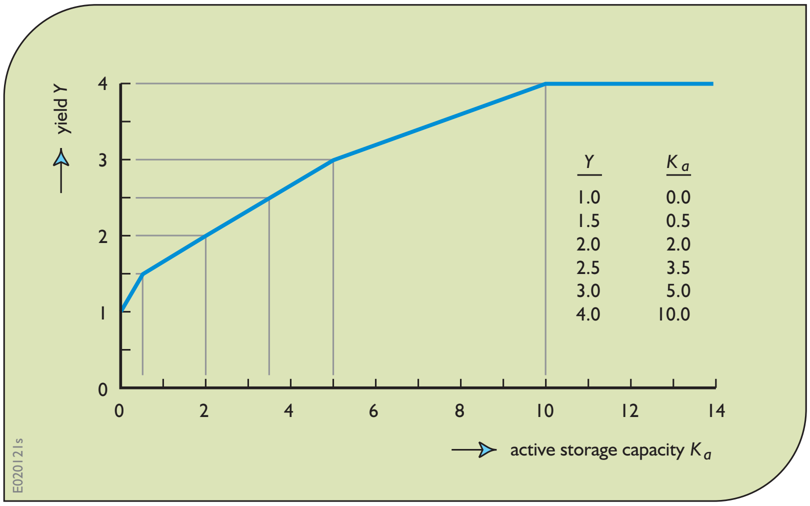



**Figure 11.14.** Storage–yield function for the sequence of flows 1, 3, 3, 5, 8, 6, 7, 2 and 1.

While the mass diagram and sequent peak procedures are relatively simple, they are not readily adaptable to reservoirs where evaporation losses and/or lake level regulation are important considerations, or to problems involving more than one reservoir. Mathematical programming (optimization) methods provide this capability. These optimization methods are based on mass-balance equations for routing flows through each reservoir. The mass-balance or continuity equations explicitly define storage volumes (and hence storage areas from which evaporation occurs) at the beginning of each period $t$.

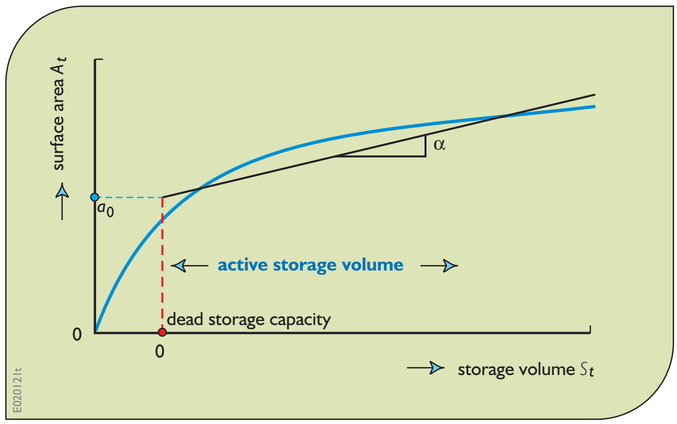



**Figure 11.15.** Storage surface area as a function of reservoir storage volume along with its linear approximation. The parameter *a* is the assumed increase in surface area associated with a unit increase in the storage volume.

### 2.2.3. Evaporation Losses

Evaporation losses, $L_t$, from lakes and reservoirs, if any, take place on their surface areas. Hence, to compute these losses, their surface areas must be estimated in each period $t$. Storage surface areas are functions of the storage volumes, $S_t$. These functions are typically concave, as shown in Figure 11.15.

In addition to the storage area–volume function, the seasonal surface water evaporation loss depths, $E_t^{\max}$, must be known. Multiplying the average surface area, $A_t$, based on the initial and final storage volumes, $S_t$ and $S_{t+1}$, by the loss depth, $E_t^{\max}$, yields the volume of evaporation loss, $L_t$, in the period $t$. The linear approximation of that loss is

$$L_t = [a_0 + a(S_t + S_{t+1})/2]E_t^{\max} \tag{11.37}$$

Letting

$$a_t = 0.5aE_t^{\max} \tag{11.38}$$

the mass-balance equation for storage volumes that include evaporation losses in each period $t$ can be approximated as:

$$(1 - a_t)S_t + Q_t - R_t - a_0E_t^{\max} = (1 + a_t)S_{t+1} \tag{11.39}$$

If Equation 11.39 is used in optimization models for identifying preliminary designs of a proposed reservoir, and if the active storage capacity turns out to be essentially zero, or just that required to provide for the fixed evaporation loss, $a_0E_t^{\max}$, then clearly any reservoir at the site is not justified. These mass-balance equations, together with any reservoir storage capacity constraints, should be removed from the model before resolving it again. This procedure is preferable to introducing 0,1 integer variables that will include the term $a_0 E_t^{\max}$ in Equation 11.39 only if the active storage volume is greater than 0 (using methods discussed in Chapter 4).

An alternative way to estimate evaporation loss that does not require a surface area–storage volume relationship, such as the one shown in Figure 11.15, is to define the storage elevation–storage volume function. Subtracting the evaporation loss depth from the initial surface elevation associated with the initial storage volume will result in an adjusted storage elevation, which in turn defines the initial storage volume after evaporation losses have been deducted. This adjusted initial volume can be used in continuity equations (Equation 11.32 or 11.36). This procedure assumes evaporation is only a function of the initial storage volume in each time period $t$. For relatively large volumes and short time periods, such an assumption is usually satisfactory.

### 2.2.4. Over and Within-Year Reservoir Storage and Yields

An alternative approach to modelling reservoirs is to separate out over-year storage and within-year storage, and to focus not on total reservoir releases, but on parts of the total releases that can be assigned specific reliabilities. These release components we call yields. To define these yields and the corresponding reservoir release rules, we divide this section into four parts. The first outlines a method for estimating the reliabilities of various constant annual minimum flows or yields. The second part discusses a modelling approach for estimating the over-year and within-year active storage capacities that are required to deliver specified annual and within-year yields, each having pre-specified reliabilities. The third and fourth parts expand this modelling approach to include multiple yields having different reliabilities, evaporation losses and the construction of reservoir operation rule curves using these flow release yields.

It will be convenient to illustrate the yield models and their solutions using a simple example of a single reservoir and two within-year periods per year. This example will be sufficient to illustrate the method that can be applied to larger systems having more within-year periods. Table 11.3 lists the nine years of available streamflow data for each within-year season at a potential reservoir site. These streamflows are used to compare the solutions of various yield models as well as to illustrate the concept of yield reliability.

| time year $y$ | within-year period $Q_{ty}$ $Q_{1y}$ | $Q_{2y}$ | annual $Q_y$ flow |
|---|---|---|---|
| 1 | 1.0 | 3.0 | 4.0 |
| 2 | 0.5 | 2.5 | 3.0 |
| 3 | 1.0 | 2.0 | 3.0 |
| 4 | 0.5 | 1.5 | 2.0 |
| 5 | 0.5 | 0.5 | 1.0 |
| 6 | 0.5 | 2.5 | 3.0 |
| 7 | 1.0 | 5.0 | 6.0 |
| 8 | 2.5 | 5.5 | 8.0 |
| 9 | 1.5 | 4.5 | 6.0 |
| **total** | **9.0** | **27.0** | **36.0** |
| average flow | 1.0 | 3.0 | 4.0 |

E020130m

**Table 11.3. Recorded unregulated historical streamflows at a reservoir site.**

### Reliability of Annual Yields

As discussed in Chapter 4, reservoirs can be operated so as to increase the dependable flow downstream. The flow that can always be depended upon at a particular site is commonly referred to as the ‘firm yield’ or ‘safe yield’. These terms imply that this is what the reservoir will always be able to provide. Of course, this may not be true. If historical flows are used to determine this yield, then the resulting yield might be better called an ‘historical yield’. Historical and firm yield are often used synonymously.

A flow yield is 100% reliable only if the sequence of flows in future years will never sum to a smaller amount than those that have occurred in the historic record. Usually one cannot guarantee this condition. Hence, associated with any historic yield is the uncertainty (a probability) that it might not always be available in the future. There are various ways of estimating this probability.

Referring to the nine-year streamflow record listed in Table 11.3, if no reservoir is built to increase the yields downstream of the reservoir site, the historic firm yield is the lowest flow on record, namely 1.0 that occurred in year 5. The reliability of this annual yield is the probability that the streamflow in any year will be greater than or equal to this value. In other words, it is the probability that this flow will be equalled or exceeded. The expected value of the exceedance probability of the lowest flow in an $n$-year record is approximately $n/(n+1)$, which for the $n = 9$ year flow record is $9/(9+1)$, or 0.90. This is based on the assumption that any future flow has an equal probability of being in any of the intervals formed by ordering the record of flows from the lowest to the highest value.

Now, rank the $n$ flows of record from the highest to the lowest. Assign the rank $m$ of 1 to the highest flow, and $n$ to the lowest flow. In general, the expected probability $p$ that any flow of rank $m$ will be equalled or exceeded in any year is approximately $m/(n+1)$. An annual yield having a probability $p$ of exceedance will be denoted as $Y_p$.

For independent events, the expected number of years until a flow of rank $m$ is equalled or exceeded is the reciprocal of its probability of exceedance $p$, namely $1/p = (n+1)/m$. The recurrence time or expected time until a failure (a flow less than that of rank $m$) is the reciprocal of the probability of failure in any year. Thus, the expected recurrence time $T_p$ associated with a flow having an expected probability $p$ of exceedance is $1/(1-p)$.

### 2.2.5. Estimation of Active Reservoir Storage Capacities for Specified Yields

#### Over-Year Storage Capacity

A reservoir with active over-year storage capacity provides a means of increasing the magnitude and/or the reliabilities of various annual yields. For example, the sequent peak algorithm defined by Equation 11.30 provides a means of estimating the reservoir storage volume capacity required to meet various 'firm' yields, $Y_{0.9}$, associated with the nine annual flows presented in Table 11.3. The same yields can be obtained from a linear optimization model that minimizes active over-year storage capacity, $K_a^o$,

$$\text{Minimize } K_a^o \tag{11.40}$$

required to satisfy the following storage continuity and capacity constraint equations involving only annual storage volumes, $S_y$, inflows, $Q_y$, yields, $Y_p$, and excess releases, $R_y$. For each year $y$:

$$S_y + Q_y - Y_p - R_y = S_{y+1} \tag{11.41}$$

$$S_y \leq K_a^o \tag{11.42}$$

Once again, if the year $y$ is the last year of record, then year $y + 1$ is assumed to equal 1. For annual yields of 3 and 4, the over-year storage requirements are 3 and 8 respectively, as can be determined just by examining the right hand column of annual flows in Table 11.3.

The over-year model, Equations 11.40 through 11.42, identifies only annual or over-year storage requirements based on specified (known) annual flows, $Q_y$, and specified constant annual yields, $Y_p$. Within-year periods $t$ requiring constant yields $y_{pt}$ that sum to the annual yield $Y_p$ may also be considered in the estimation of the required over-year and within-year or total active storage capacity, $K_a$. Any distribution of the over-year yield within the year that differs from the distribution of the within-year inflows may require additional active reservoir storage capacity. This additional capacity is called the within-year storage capacity.

The sequent peak method, Equation 11.30, can be used to obtain the total over-year and within-year active storage capacity requirements for specified within-year period yields, $y_{pt}$. Alternatively a linear programming model can be developed to obtain the same information along with associated reservoir storage volumes. The objective is to find the minimum total active storage capacity, $K_a$, subject to storage volume continuity and capacity constraints for every within-year period of every year. This model is defined as:

$$\text{minimize } K_a \tag{11.43}$$

subject to

$$S_{ty} + Q_{ty} - y_{pt} - R_{ty} = S_{t+1,\,y} \quad \forall t, y \tag{11.44}$$

$$S_{ty} \leq K_a \quad \forall t, y \tag{11.45}$$

In Equation 11.44, if $t$ is the final period $T$ in year $y$, the next period $T + 1 = 1$ in year $y + 1$, or year 1 if $y$ is the last year of record, $n$.

The difference in the active capacities resulting from these two models, Equations 11.40 through 11.42, and Equations 11.43 through 11.45, is the within-year storage requirement, $K_a^w$.

Table 11.4 shows some results from solving both of the above models. The over-year storage capacity requirements, $K_a^0$, are obtained from Equations 11.40 through 11.42. The combined over-year and within-year capacities, $K_a$, are obtained from solving Equations 11.43 through 11.45. The difference between the over-year storage capacity, $K_a^0$, required to meet only the annual yields and the total capacity, $K_a$, required to meet each specified within-year yield distribution of those annual yields is the within-year active storage capacity $K_a^w$.

Clearly, the number of continuity and reservoir capacity constraints in the combined over-year and within-year model (Equations 11.43–11.45) can become very large when the number of years $n$ and within-year periods $T$ are large. Each reservoir site in the river system will require $2nT$ continuity and capacity constraints. Not all these constraints are necessary, however. It is only a subset of the sequence of flows within the total record of flows that generally determines the required active storage capacity $K_a$ of a reservoir. This is called the *critical period*. This critical period is often used in engineering studies to estimate the historical yield of any particular reservoir or system of reservoirs.

Even though the severity of future droughts is unknown, many planners accept the traditional practice of using the historical critical drought period for reservoir design and operation studies. They assume events that have happened in the past could happen again in the future. In some parts of the world, notably those countries in the lower portions of the southern hemisphere, historical records are continually proven to be unreliable indicators of future hydrological conditions. In these regions especially, synthetically generated flows based on statistical methods are more acceptable as a basis for yield estimation than seems to be the case elsewhere.

### Over and Within-Year Storage Capacity

To begin the development of a smaller, but more approximate, model, consider each combined over-year and within-year storage reservoir to consist of two separate reservoirs in series (Figure 11.16). The upper reservoir is the over-year storage reservoir, whose capacity required for an annual yield is determined by an over-year model, such as Equations 11.40 through 11.42. The purpose of the 'downstream' within-year reservoir is to distribute as desired in each within-year period $t$ the annual yield produced by the 'upstream' over-year reservoir. Within-year storage capacity would not be needed if the distribution

E020130n

| annual yield $y^{0.9}$ | within-year yields $t = 1$ | $t = 2$ | required active storage volume capacity: within-year $K_a^w$ | over-year $K_a^o$ | total $K_a$ |
|---|---|---|---|---|---|
| 3 | 0 | 3 | 1.0 | 3.0 | 4.0 |
| | 1 | 2 | 0.5 | 3.0 | 3.5 |
| | 2 | 1 | 1.5 | 3.0 | 4.5 |
| | 3 | 0 | 2.5 | 3.0 | 5.5 |
| 4 | 0 | 4 | 1.0 | 8.0 | 9.0 |
| | 1 | 3 | 0.0 | 8.0 | 8.0 |
| | 2 | 2 | 1.0 | 8.0 | 9.0 |
| | 3 | 1 | 2.0 | 8.0 | 10.0 |
| | 4 | 0 | 3.0 | 8.0 | 11.0 |

**Table 11.4.** Active storage requirements for various within-year yields.

of the average inflows into the upper over-year reservoir exactly coincided with the desired distribution of within-year yields. If they do not, within-year storage may be required. The two separate reservoir capacities summed together will be an approximation of the total active reservoir storage requirement needed to provide those desired within-year period yields.

Assume the annual yield produced and released by the over-year reservoir is distributed in each of the within-year periods in the same ratio as the average within-year inflows divided by the total average annual inflow. Let the ratio of the average period $t$ inflow divided by the total annual inflow be $\beta_t$. The general within year model is to find the minimum within-year storage capacity, $K_a^w$, subject to within-year storage volume continuity and capacity constraints.

$$\text{Minimize } K_a^w \tag{11.46}$$

subject to

$$s_t + \beta_t Y_p - y_{pt} = s_{t+1} \quad \forall t \quad T+1 = 1 \tag{11.47}$$

$$s_t \leq K_a^w \quad \forall t \tag{11.48}$$

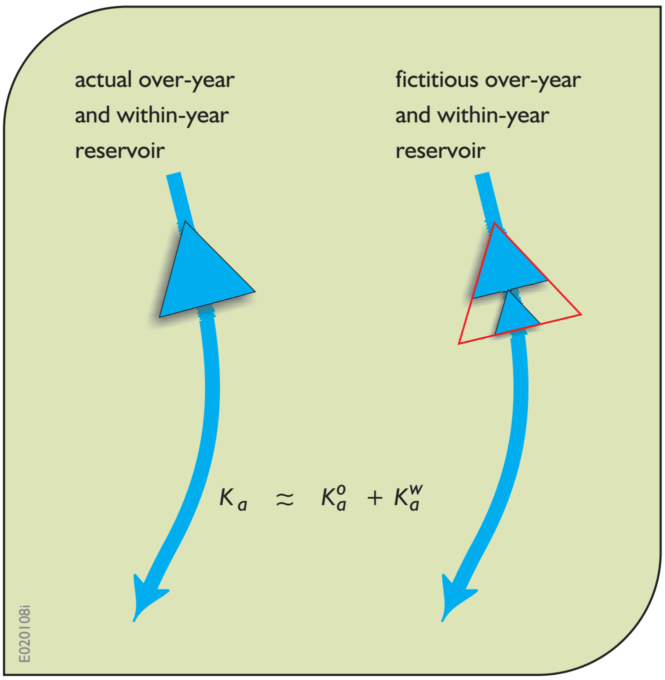



**Figure 11.16.** Approximating a combined over-year and within-year reservoir as two separate reservoirs, one for creating annual yields, the other for distributing them as desired in the within-year periods.

Since the sum of $\beta_t$ over all within-year periods $t$ is 1, the model guarantees that the sum of the unknown within-year yields, $y_{pt}$, equals the annual yield, $Y_{pt}$.

The over-year model, Equations 11.40 through 11.42, and within-year model, Equations 11.46 through 11.48, can be combined into a single model for an $n$-year sequence of flows:

$$\text{Minimize } K_a \tag{11.49}$$

subject to:

$$S_y + Q_y - Y_p - R_y = S_{y+1} \quad \forall y \quad \text{if } y = n, y+1 = 1 \tag{11.50}$$

$$S_y \leq K_a^o \quad \forall y \tag{11.51}$$

$$s_t + \beta_t Y_p - y_{pt} = s_{t+1} \quad \forall t \quad \text{if } t = T, T+1 = 1 \tag{11.52}$$

$$s_t \leq K_a^w \quad \forall t \tag{11.53}$$

$$\sum_t y_{pt} = Y_p \tag{11.54}$$

$$K_a \geq K_a^o + K_a^w \tag{11.55}$$

Constraint 11.54 is not required due to Equation 11.52, but is included here to make it clear that the sum of within-year yields will equal the over-year yield, and that such a constraint will be required for each yield of reliability $p$ if multiple yields of different reliabilities are included in the model. In addition, constraint Equation 11.53 can be combined with Equation 11.55, saving a constraint. If this is done, the combined model contains $2n + 2T + 1$ constraints, compared to the more accurate model, Equations 11.43 through 11.45, that contains $2nT$ constraints.

If the fractions $\beta_t$ are based on the ratios of the average within-year inflow divided by average annual inflow in the two within-year periods shown in Table 11.3, then 0.25 of the total annual yield flows into the fictitious within-year reservoir in period $t = 1$, and 0.75 of the total annual yield flows into the reservoir in period $t = 2$. Suppose the two desired within-year yields are to be 3 and 0 for periods 1 and 2 respectively. The total annual yield, $Y_{0.9}$, is 3. Assuming the natural distribution of this annual yield of 3 in period 1 is 0.25 $Y_{0.9} = 0.75$, and in period 2 it is 0.75 $Y_{0.9} = 2.25$, the within-year storage required to redistribute these yields of 0.75 and 2.25 to become 3 and 0, respectively, is $K_a^w = 2.25$. From Table 11.4, we can see an annual yield of 3 requires an over-year storage capacity of 3. Thus, the estimated total storage capacity required to provide yields of 3 and 0 in periods 1 and 2 is the over-year capacity of 3 plus the within-year capacity of 2.25 equalling 5.25. This compares with 3 plus 2.5 of actual within-year capacity required for a total of 5.50, as indicated in Table 11.4.

There are ways to reduce the number of over-year constraints without changing the solution of the over-year model. Sequences of years whose annual inflow values equal or exceed the desired annual yield can be combined into one constraint. If the yield is an unknown variable, then the mean annual inflow can be used since it is the upper limit of the annual yield. For example in Table 11.3 note that the last three years and the first year have flows equal or greater than 4, the mean annual inflow. Thus,these four successive years can be combined into a single continuity equation:

$$S_7 + Q_7 + Q_8 + Q_9 + Q_1 - 4Y_p - R_7 = S_2 \qquad (11.56)$$

This saves a total of three over-year continuity constraints and three over-year capacity constraints. Note that the excess release, $R_7$, represents the excess release in all four periods. Furthermore, not all reservoir capacity constraint Equation 11.51 are needed, since the initial storage volumes in the years following low flows will probably be less than the over-year capacity.

There are many ways to modify and extend this yield model to include other objectives, fixed ratios of the unknown annual yield for each within-year period, and even multiple yields having different exceedance probabilities $p$.

The number of over-year periods being modelled compared to the number of years of flow records determines the highest exceedance probability or reliability a yield can have, e.g. 9/10 or 0.9 in the nine-year example used here. If yields having lower reliabilities are desired, such as a yield with a reliability of 0.80, then the yield variable $Y_P$ ($Y_{0.90}$) can be omitted from Equation 11.50 in the critical year that determines the required over-year capacity for a 0.90 reliable yield. (Since some outflow might be expected, even if it is less than the 90% reliable yield, the outflow could be forced to equal the inflow for that year.) If a 0.70 reliable yield is desired, then the yield variables in the two most critical years can be omitted from Equation 11.50, and so on.

The number of years of yield failure determines the estimated reliability of each yield. An annual yield that fails in $f$ out of $n$ years has an estimated probability $(n-f)/(n+1)$ of being equalled or exceeded in any future year. Once the desired reliability of a yield is known, the modeller must select the appropriate failure years and specify the permissible extent of failure in those $f$ failure years.

To consider different yield reliabilities $p$, let the parameter $\alpha_y^p$ be a specified value between 0 and 1 that indicates the extent of a failure in year $y$ associated with an annual yield having a reliability of $p$. When $\alpha_y^p$ is 1 there is no failure, and when it is less than 1 there is a failure, but a proportion of the yield $Y_p$ equal to $\alpha_y^p$ is released. Its value is in part dependent on the consequences of failure and on the ability to forecast when a failure may occur and to adjust the reservoir operating policy accordingly.

The over-year storage continuity constraints for $n$ years can now be written in a form appropriate for identifying any single annual yield $Y_p$ having an exceedance probability $p$:

$$S_y + Q_y - \alpha_y^p Y_p - R_y = S_{y+1} \quad \forall y$$
$$\text{if } y = n, y+1 = 1 \qquad (11.57)$$

When writing Equation 11.57, the failure year or years should be selected from among those in which permitting a failure decreases the required reservoir capacity $K_a$. If a failure year is selected that has an excess release, no reduction in the required active storage capacity will result and the reliability of the yield may be higher than intended.

The critical year or years that determine the required active storage volume capacity may be dependent on the yield itself. Consider, for example, the seven-year sequence of annual flows (4, 3, 3, 2, 8, 1, 7) whose mean is 4. If a yield of 2 is desired in each of the seven years, the critical year requiring reservoir capacity is year 6. If a yield of 4 is desired (again assuming no losses), the critical years are years 2 through 4. The streamflows and yields in these critical years determine the required over-year storage capacity. The failure years, if any, must be selected from within the critical low flow periods for the desired yield.

When the magnitudes of the yields are unknown, some trial-and-error solutions may be necessary to ensure that any failure years are within the critical period of years for the associated yields. To ensure a wider range of applicable yield magnitudes, the year having the lowest flow within the critical period should be selected as the failure year if only one failure year is selected. Even though the actual failure year may follow that year, the resulting required reservoir storage volume capacity will be the same.

### Multiple Yields and Evaporation Losses

The yield models developed so far define only single annual and within-year yields. Incremental secondary yields having lower reliabilities can also be included in the model. Referring to the nine-year streamflow record in Table 11.3, assume that two yields are desired, one 90% reliable and the other 70% reliable. Let $Y_{0.9}$ and $Y_{0.7}$ represent those annual yields having reliabilities of 0.9 and 0.7, respectively. The incremental secondary yield $Y_{0.7}$ represents the amount in addition to $Y_{0.9}$ that is only 70% reliable. Assume that the problem is one of estimating the appropriate values of $Y_{0.9}$ and $Y_{0.7}$, their respective within-year allocations $y_{pt}$, and the total active reservoir capacity $K_a$ that maximizes some function of these yield and capacity variables.

In this case, the over-year and within-year continuity constraints can be written

$$S_y + Q_y - Y_{0.9} - \alpha_y^{0.7} Y_{0.7} - R_y = S_{y+1} \quad \forall y$$
$$\text{if } y = n,\ y+1 = 1 \tag{11.58}$$

$$s_t + \beta_t (Y_{0.9} + Y_{0.7}) - y_{0.9,t} - y_{0.7,t} = s_{t+1} \quad \forall t$$
$$\text{if } t = T,\ T+1 = 1 \tag{11.59}$$

Now an additional constraint is needed to ensure that all within-year yields of a reliability $p$ sum to the annual yield of the same reliability. Selecting the 90% reliable yield,

$$\sum_t y_{0.9,t} = Y_{0.9} \tag{11.60}$$

Evaporation losses must be based on an expected storage volume in each period and year, since the actual storage volumes are not identified using these yield models. The approximate storage volume in any period $t$ in year $y$ can be defined as the initial over-year volume $S_y$, plus the estimated average within-year volume $(s_t + s_{t+1})/2$. Substituting this storage volume into Equation 11.37 (see also Figure 11.15) results in an estimated evaporation loss $L_{yt}$:

$$L_{yt} = [a_0 + a(S_y + (s_t + s_{t+1})/2)]E_t^{\max} \tag{11.61}$$

Summing $L_{yt}$ over all within-year periods $t$ defines the estimated annual evaporation loss, $E_y$:

$$E_y = \sum_t [a_0 + a(S_y + (s_t + s_{t+1})/2)]E_t^{\max} \tag{11.62}$$

This annual evaporation loss applies, of course, only when there is a non-zero active storage capacity requirement. These annual evaporation losses can be included in the over-year continuity constraints, such as Equation 11.58. If they are, the assumption is being made that their within-year distribution will be defined by the fractions $\beta_t$. This may not be realistic. If it is not, an alternative would be to include the average within-year period losses, $L_t$, in the within-year constraints.

The average within-year period loss, $L_t$, can be defined as the sum of each loss $L_{yt}$ defined by Equation 11.61 over all years $y$ divided by the total number of years, $n$:

$$L_t = \sum_y^n [a_0 + a(S_y + (s_t + s_{t+1})/2)]E_t^{\max}/n \tag{11.63}$$

This average within-year period loss, $L_t$, can be added to the within-year's highest reliability yield, $y_{pt}$, forcing greater total annual yields of all reliabilities to meet corresponding total within-year yield values. Hence, combining Equations 11.60 and 11.61, for $p$ equal to 0.9 in the example,

$$Y_p = \sum_t \left\{ y_{pt} + \sum_y^n [a_0 + a(S_y + (s_t + s_{t+1})/2)]E_t^{\max}/n \right\} \tag{11.64}$$

Since actual reservoir storage volumes in each period $t$ of each year $y$ are not identified in this model, system performance measures that are functions of those storage volumes, such as hydroelectric energy or reservoir recreation, are only approximate. Thus, as with any of these screening models, any set of solutions should be evaluated and further improved using more precise simulation methods.

Simulation methods require reservoir operating rules. The information provided by the solution of the yield model can help in defining a reservoir operating policy for such simulation studies.

### Reservoir Operation Rules

Reservoir operation rules, as discussed in Chapters 4, 7 and 8, are guides for those responsible for reservoir operation. They apply to reservoirs being operated in a steady-state condition, i.e. not filling up immediately after construction or being operated to meet a set of new and temporary objectives. There are several types of rules, but each indicates the desired or required reservoir release or storage volume at any particular time of year. Some rules identify storage volume targets (rule curves) that the operator is to maintain, if possible, and others identify storage zones, each associated with a particular release policy. This latter type of rule can be developed from the solution of the yield model.

To construct an operation rule that identifies storage zones, each having a specific release policy, the values of the dead and flood storage capacities, $K_D$ and $K_f$, are needed together with the over-year storage capacity, $K_a^o$, and within-year storage volumes, $s_t$, in each period $t$. Since both $K_a^o$ and all $s_t$ derived from the yield model are for all yields, $Y_p$, being considered, it is necessary to determine the over-year capacities and within-year storage volumes required to provide each separate within-year yield, $y_{pt}$. Plotting the curves defined by the respective over-year capacity plus the within-year storage volume $(K_a^o + s_t)$ in each within-year period $t$ will define a zone of storage whose yield releases $y_{pt}$ from that zone should have a reliability of at least $p$.

For example, assume again a nine-year flow record and ten within-year periods. Of interest are the within-year yields, $y_{0.9,t}$ and $y_{0.7,t}$, having reliabilities of 0.9 and 0.7. The first step is to compute the over-year storage capacity requirement, $K_a^o$, and the within-year storage volumes, $s_t$, for just the yields $y_{0.9,t}$. The sum of these values, $K_a^o + s_t$, in each period $t$ can be plotted as illustrated in Figure 11.17.

The sum of the over-year capacity and within-year volume $K_a^o + s_t$ in each period $t$ defines the zone of active storage volumes for each period $t$ required to supply the within-year yields $y_{0.9,t}$. If the storage volume is in the shaded zone shown in Figure 11.17, only the yields $y_{0.9,t}$ should be released. The reliability of these yields, when simulated, should be about 0.9. If at any time $t$ the actual reservoir storage volume is within this zone, then reservoir releases should not exceed those required to meet the yield $y_{0.9,t}$ if the reliability of this yield is to be maintained.

The next step is to solve the yield model for both yields $Y_{0.9}$ and $Y_{0.7}$. The resulting sum of over-year storage capacity and within-year storage volumes can be plotted over the first zone, as shown in Figure 11.18.

If at any time $t$ the actual storage volume is in the second, lighter-shaded, zone in Figure 11.18, the release should be the sum of both the most reliable yield, $y_{0.9,t}$ and the incremental secondary yield $y_{0.7,t}$. If only these releases are made, then the probability of being in that zone, when simulated, should be about 0.7. If the actual storage volume is greater than the total required over-year storage capacity $K_a^o$ plus the within-year volume $s_t$, the non-shaded zone in Figure 11.18, then a release can

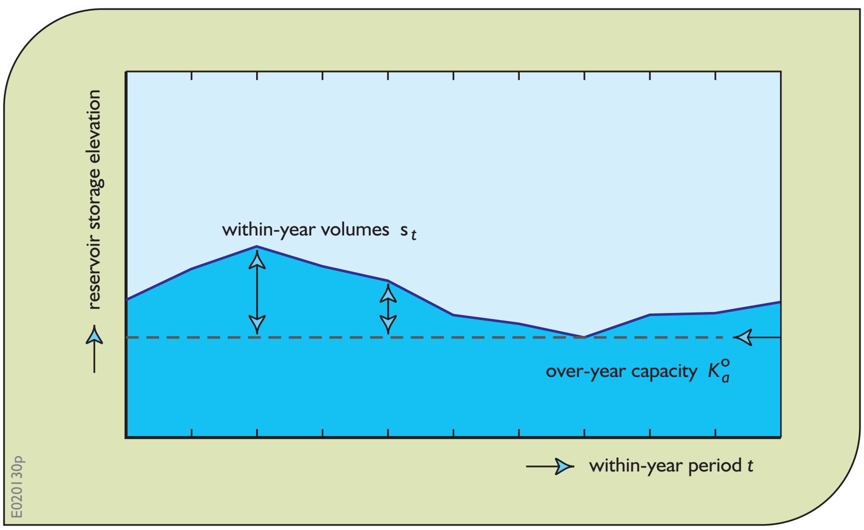



**Figure 11.17.** Reservoir release rule showing the identification of the most reliable release zone associated with the within-year yields $y_{0.9,t}$.

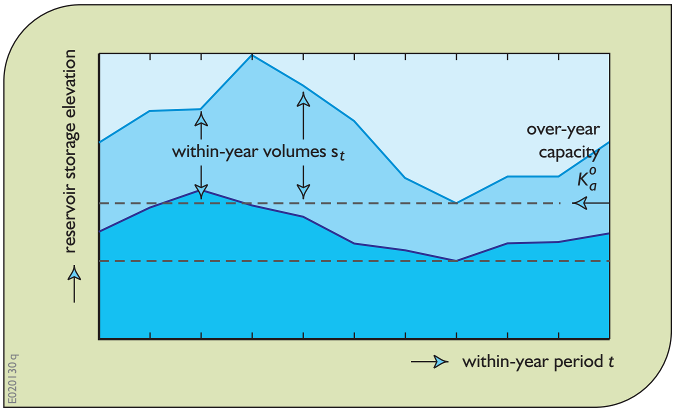



**Figure 11.18.** Reservoir release rule showing the identification of the second most reliable release zone associated with the total within-year yields $y_{0.9,t} + y_{0.7,t}$.

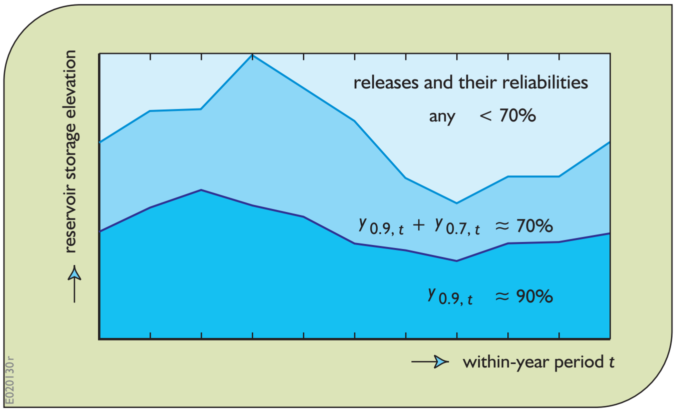



**Figure 11.19.** Reservoir release rule defined by the yield model.

be made to satisfy any down-stream demand. Converting storage volume to elevation, this release policy is summarized in Figure 11.19.

These yield models focus only on the active storage capacity requirements. They can be a part of a model that includes flood storage requirements as well (as discussed later in this chapter). If the actual storage volume is within the flood control zone in the flood season, releases should be made to reduce the actual storage to a volume no greater than the total capacity less the flood storage capacity.

Once again, reservoir rules developed from simplified models such as this yield model are only guides, and once developed they should be simulated, evaluated and refined prior to their actual adoption.

## 2.3. Wetlands and Swamps

Wetlands and swamps are important elements of the water resources system. They provide regulating functions with respect to both quantity (e.g. flood attenuation) and quality (e.g. self-purification). Most wetlands support valuable ecological systems. Appendix A (Section 4) describes the natural processes involved in wetlands.

From a quantitative point of view, wetlands are comparable with (shallow) lakes. Lakes are usually located on streams or rivers, while most wetlands are separate elements connected to a stream or river only in wet periods, either by flooding from the stream or river or by draining to it. Some wetlands are part of river systems. Examples are the Sudd Swamps in the Nile River, the Pantanal in the Paraguay River in South America and the marshlands in southern Iraq at the confluence of the Euphrates and Tigris Rivers. In these cases wetlands can be modelled as lakes or reservoirs, as described in the previous section. If they are separate from the river system, wetlands can be included as indicated in Figure 11.20. The lateral flow $Q_3(t)$ can be bi-directional.

## 2.4. Water Quality and Ecology

River basin models focusing on water quantities are mostly used to investigate whether sufficient water is available to satisfy the various use functions (off-stream and in-stream), and to identify measures to match supply and demand. The core of most river basin models consists of keeping track of the water balance of the whole river basin. The analysis of water quality and ecology is mostly done 'off-line', using another model for a specific part of the system, e.g. a river stretch, reservoir or groundwater system. These models will be described in Chapter 12.

Such an off-line approach makes sense. There is little feedback from water quality to quantity (except in cases where minimum flows are required to maintain a minimum water quality level). Using separate water quality models for parts of the system makes it possible to include more temporal and spatial detail and to include more complex water quality processes. Environmental flow requirements can be included in river basin models by defining specific flow regime demands (quantity, velocity, dynamics and the like) at certain locations in the river basin.

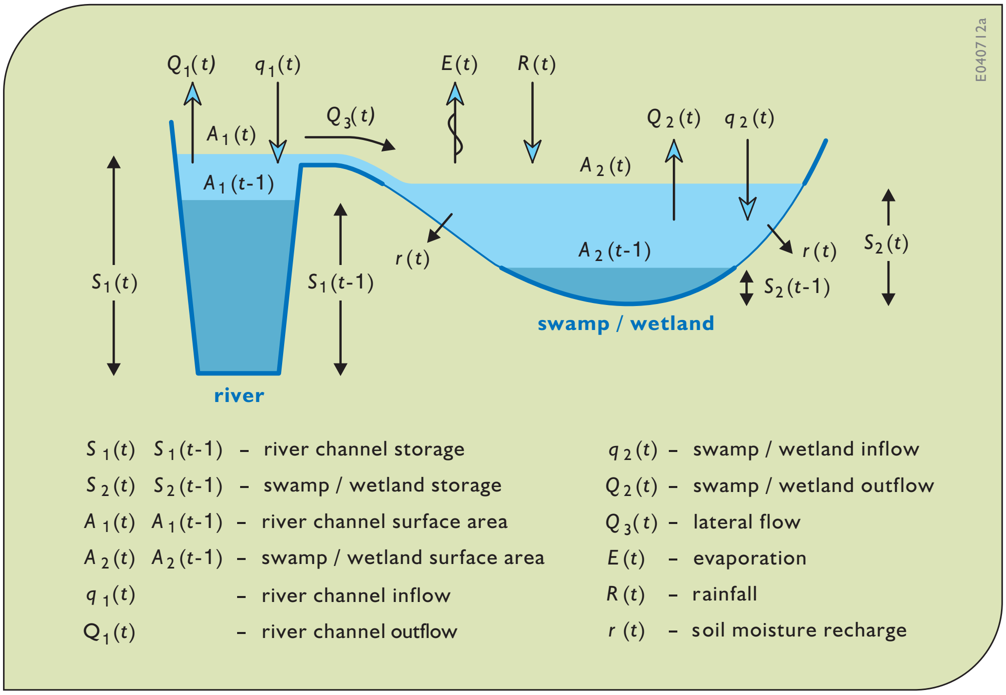



**Figure 11.20.** Schematization of the link between wetlands and a river system.

# 3. Modelling the Socio-Economic Functions In a River Basin

## 3.1. Withdrawals and Diversions

Major demands for the withdrawal of water include those for domestic or municipal uses, industrial uses (including cooling water) and agricultural uses (including irrigation). These uses generally require the withdrawal of water from a river system or other surface or groundwater body. The water withdrawn may be only partially consumed, and that which is not consumed may be returned to the river system, perhaps at a different site, at a later time period, and containing different concentrations of constituents.

Water can also be allocated to instream uses that alter the distribution of flows in time and space. Such uses include:

- reservoir storage, possibly for recreational use as well as for water supply
- flow augmentation, possibly for water quality control, navigation or ecological benefits
- hydroelectric power production.

The instream uses may complement or compete with each other or with various off-stream municipal, industrial and agricultural demands. One purpose of developing management models of river basin systems is to help derive policies that will best serve these multiple uses, or at least identify the tradeoffs among the multiple purposes and objectives.

The allocated flow $q_t^s$ to a particular use at site $s$ in period $t$ must be no greater than the total flow available, $Q_t^s$, at that site and in that period:

$$q_t^s \leq Q_t^s \tag{11.65}$$

The quantity of water that any particular user expects to receive in each particular period is termed the *target allocation*. Given an annual (known or unknown) target allocation $T^s$ at site $s$, some (usually known) fraction, $f_t^s$, of that annual target allocation will be expected in each within-year period $t$. If the actual allocation, $q_t^s$, is less than the target allocation, $f_t^s T^s$, there will be a deficit, $D_t^s$. If the allocation is greater than the target allocation, there will be an excess, $E_t^s$. Hence, to define those unknown variables, the following constraint equation can be written for each applicable period $t$:

$$q_t^s = f_t^s T^s - D_t^s + E_t^s \tag{11.66}$$

Even though allowed, one would not expect a solution to contain non-zero values for both $D_t^s$ and $E_t^s$.

Whether or not any deficit or excess allocation should be allowed at any demand site $s$ depends on the quantity of water available and the losses or penalties associated with deficit or excess allocations to that site. At sites where the benefits derived in each period are independent of the allocations in other periods, the losses associated with deficits and the losses or benefits associated with excesses can be defined in each period $t$ (Chapter 10). For example, the benefits derived from the allocation of water for hydropower production in period $t$ will in some cases be essentially independent of previous allocations.

For any use in which the benefits are dependent on a sequence of allocations, such as at irrigation sites, the benefits may be based on the annual (or growing season) target water allocations $T^s$ and their within season distributions, $f_t^s T^s$. In these cases one can define the benefits from those water uses as functions of the unknown season or annual targets, $T^s$, where the allocated flows $q_t^s$ must be no less than the specified fraction of that unknown target;

$$q_t^s \geq f_t^s T^s \quad \text{for all relevant } t \qquad (11.67)$$

If, for any reason, an allocation $q_t^s$ must be zero, then clearly from Equation 11.67 the annual or growing season target allocation $T^s$ would be zero, and presumably so would be the benefits associated with that target value.

Water stored in reservoirs can often be used to augment downstream flows for instream uses such as recreation, navigation and water quality control. During natural low-flow periods in the dry season, it is not only the increased volume but also the lower temperature of the augmented flows that may provide the only means of maintaining certain species of fish and other aquatic life. Dilution of wastewater or runoff from non-point sources may be another potential benefit from flow augmentation. These and other factors related to water quality management are discussed in greater detail in Chapter 12.

The benefits derived from navigation on a potentially navigable portion of a river system can usually be expressed as a function of the stage or depth of water in various periods. Assuming known stream or river flow-stage relationships at various sites in the river, a possible constraint might require at least a minimum acceptable depth, and hence flow, for those sites.

## 3.2. Domestic, Municipal and Industrial Water Demand

Domestic, municipal and industrial (DMI) water demands are typically based on projections of population and socio-economic activities. The domestic demand is the household demand. The municipal demand covers the water requirements of the commercial sector (shops, department stores, hotels and so on) and for the services sector (hospitals, offices, schools and others). In most river basins, the DMI demand will be small compared to the demand for irrigation. However, given the importance of this demand for human health and economic developments, it is often given first priority. The preferred source for drinking water (domestic and municipal) is often groundwater which requires only limited treatment. When sufficient groundwater of good quality is not available, surface water will be used, in most cases in combinati4on with advanced treatment facilities.

About 80% to 90% of the water abstracted for DMI will usually be returned to the system as wastewater. About 95% of cooling water is also typically returned to the basin. In river basin studies, the total demand has to be taken into account, as this amount is actually withdrawn from the system. The water not used will be returned to the system, possibly at another location, in a future time period and containing more pollutants because of its use.

Predicting DMI demands can be difficult. A common approach is to use the demand predictions of the public water supply authorities in the basin. They base their predictions on estimates of population growth, growth in demand per capita and so on, often using statistical trends. As these authorities have a good knowledge of what is going on in their area, their estimates are often the most reasonable ones to use, although they often tend to overestimate the demand.

In many cases these predictions are not available or not sufficiently reliable, and it is then usually necessary to collect basic information on population and socio-economic activities and expected growth rates. Domestic use can be estimated on the basis of water consumption per capita and coverage rates, taking into account differences in social strata. Municipal demand is often assumed to be a percentage of the domestic water demand (usually in the range of 15–35%). Industrial water demand is the most difficult to estimate, as it depends on the type of industry and the production processes being used. Not only is it difficult to get all the data needed to make accurate predictions, but future economic and technological developments are very uncertain. In most cases estimates are made on the basis of statistical projections or of water demand per industrial employee (for which data are easier to get), depending on the type of industry.

Once a total demand for DMI has been determined, projections have to be made for the losses in the system. 'Unaccounted for' water losses depend on the condition and maintenance of the distribution system and should be taken into account. In some systems such losses can exceed 40%.

### 3.3. Agricultural Water Demand

Most agricultural areas in the world depend on rain as the primary source of water. During dry periods, additional water may be supplied through irrigation systems. In semi-arid and arid countries, irrigation is an absolute requirement to enable agricultural production. In those cases, agricultural water is often the dominant demand category. The irrigation demand in Egypt accounts for 90% of the total net water demand, and in the western United States for about 75% of the total.

Irrigation water demand depends on a large number of factors, which can be grouped under four headings:

- hydro-meteorological conditions
- types of crops
- soil types
- irrigation practices and water use efficiency.

Crops transpire as they grow. At the same time, there is evaporation from the soil surface. The combined quantity of water used under conditions of optimum availability is known as 'consumptive use' or 'evapotranspiration' ($ET$). The potential evapotranspiration, $E_p$, is estimated from the following equation:

$$E_p = c_p ET_0 \tag{11.68}$$

in which:

$E_p$ = potential crop evapotranspiration (mm per time unit)

$c_p$ = crop coefficient

$ET_0$ = reference crop evapotranspiration (mm per time unit)

The reference crop evapotranspiration, $ET_0$, can be computed using a standard FAO method (FAO, 1998) based on sunshine, temperature, humidity and wind speed. The crop coefficient, $c_p$, is related to the type of crop and the growth stage.

Based on crop water demand ($E_p$), the crop irrigation demand ($C_d = E_p -$ effective rainfall), field irrigation demand ($F_d = C_d +$ field losses) and project irrigation demand ($P_d = F_d +$ distribution losses) can be calculated.

The calculation of irrigation demand is complex as farmers use many different cropping patterns and cropping calendars. Several computer programs (e.g. the FAO's CROPWAT) are available to support that calculation, and many river basin simulation packages (see Section 4.3) have modules that estimate the irrigation demand.

### 3.4. Hydroelectric Power Production

Hydropower plants, Figure 11.21, convert the energy from the flowing water into mechanical and then electrical energy. These plants, containing turbines and generators, are installed either in or adjacent to dams. Pipelines (penstocks) carry water under pressure from the reservoir to the powerhouse. Power transmission systems transport energy from the powerhouse to where it is needed.

The principal advantages of hydropower are the absence of polluting emissions during operation, its capability to respond quickly to changing utility load demands, and its relatively low operating costs. Disadvantages can include high initial capital cost, potential site-specific impacts and cumulative environmental ones. The potential environmental impacts of hydropower projects include altered flow regimes below storage reservoirs, water quality degradation, mortality of fish that pass through hydroelectric turbines, blockage of fish migration and flooding of terrestrial ecosystems by impoundments. Proper design and operation of hydropower projects can mitigate some of these impacts (see also WCD, 2000). Hydroelectric projects can also provide beneficial effects such as recreation in reservoirs or in tailwaters below dams.

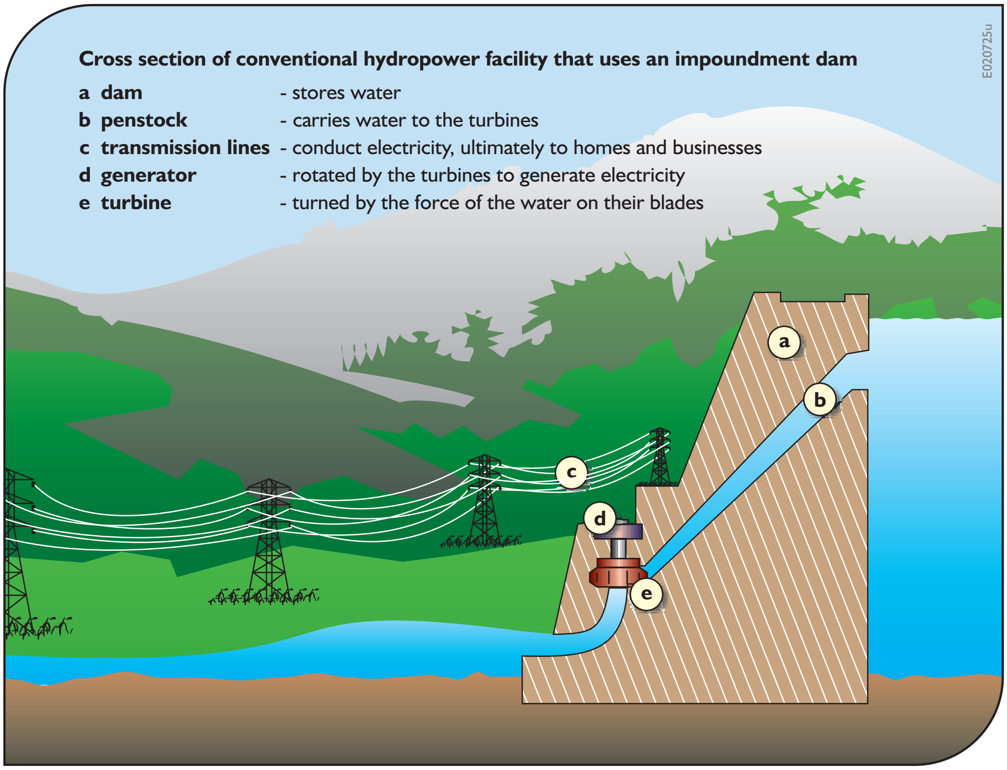



**Figure 11.21.** Hydropower plant and system components.

Hydropower technology can be grouped into two types: *conventional* and *pumped storage*. Conventional plants use the available water from a river, stream, canal system or reservoir to produce electrical energy. In conventional multipurpose reservoirs and run-of-river systems, hydropower production is just one of many competing purposes for which water resources may be used. Competing water uses can include irrigation, flood control, navigation, and municipal and industrial water supply. Pumped storage plants pump the water, usually through a reversible turbine, from a lower to an upper reservoir. While pumped storage facilities are net energy consumers, they are income producers. They are valued by the utilities because they can be rapidly brought online to operate in a peak power production mode when energy demand and prices are highest. The pumping to replenish the upper reservoir is performed during off-peak hours when electricity costs are lowest. This process benefits the utility by increasing the load factor and reducing the cycling of its base-load units. In most cases, pumped storage plants run a full cycle every twenty-four hours (USDOE, 2002).

Run-of-river plants use the natural flow of the river and produce relatively little change in the stream channel and streamflow, in contrast to a peaking plant that impounds water when the energy is needed. A storage project extensively impounds and stores water during high-flow periods to augment the water available during low-flow periods, allowing the flow releases and power production to be more constant. Many projects combine the modes.

The power capacity of a hydropower plant is primarily the function of the hydraulic head and the flow rate through the turbines. The hydraulic head is the elevation difference the water falls in passing through the plant or to the tailwater, whichever elevation difference is less. Project design may concentrate on either of these variables or both, and on the hydropower plant's installed capacity.

The production of hydroelectric energy during any period at any particular reservoir site is dependent on the installed plant capacity; the flow through the turbines; the average effective productive storage head; the number of hours in the period; the plant factor (the fraction of time that energy is produced); and a constant for converting the product of flow, head and plant efficiency to electrical energy. The kilowatt-hours of energy, $KWH_t$, produced in period $t$ is proportional to the product of the plant efficiency, $e$, the productive storage head, $H_t$, and the flow, $q_t$, through the turbines.

A cubic metre of water, weighing $10^3$ kg, falling a distance of one metre, acquires $9.81 \times 10^3$ joules (Newton-metres) of kinetic energy. The energy generated in one second equals the watts (joules per second) of power produced. Hence, an average flow of $q_t$ cubic metres per second falling a height of $H_t$ *metres* in period $t$ yields $9.81 \times 10^3 q_t H_t$ watts or $9.81\ q_t H_t$ kilowatts of power. Multiplying by the number of hours in period $t$ yields the kilowatt-hours of energy produced from an average flow rate of $q_t$. The total kilowatt-hours of energy, $KWH_t$, produced in period $t$, assuming 100% efficiency in conversion of potential to electrical energy, is

$$\begin{aligned} KWH_t &= 9.81 q_t H_t (\text{seconds in period } t)/ \\ &\quad (\text{seconds per hour}) \\ &= 9.81 q_t H_t (\text{seconds in period } t)/3600 \qquad (11.69) \end{aligned}$$

Since the total flow, $Q_t^T$ through the turbines in period $t$ equals the average flow rate $q_t$ times the number of seconds in the period, the total kilowatt-hours of energy produced in period $t$, given a plant efficiency of $e$, equals

$$\begin{aligned} KWH_t &= 9.81 Q_t^T H_t e/3600 \\ &= 0.002725 Q_t^T H_t e \qquad (11.70) \end{aligned}$$

The energy required for pumped storage, where instead of producing energy the turbines are used to pump water up to a higher level, is

$$KWH_t = 0.002725 Q_t^T H_t / e \qquad (11.71)$$

For Equations 11.70 and 11.71, $Q_t^T$ is in cubic metres and $H_t$ is in metres. The storage head, $H_t$, is the vertical distance between the water surface elevation in the lake or reservoir that is the source of the flow through the turbines and the maximum of either the turbine elevation or the downstream discharge elevation. In variable head reservoirs, storage heads are functions of storage volumes. In optimization models for capacity planning these heads and the turbine flows are among the unknown variables. The energy produced is proportional to the product of these two unknown variables. This results in non-separable functions in equations that must be written at each hydroelectric site for each time period $t$.

A number of ways have been developed to convert these non-separable energy production functions to separable ones for use in linear optimization models for estimating design and operating policy variable values. These methods inevitably increase the model complexity and the number of its variables and constraints. For a preliminary screening of hydropower capacities prior to a more detailed analysis (e.g. using simulation or other non-linear or discrete dynamic programming methods) one can:

1. Solve the model using both optimistic and pessimistic assumed fixed head values.
2. Compare the actual derived heads with the assumed ones and adjust the assumed heads.
3. Resolve the model.
4. Compare the capacity values.

From this iterative process one should be able to identify the range of hydropower capacities that can then be further refined using simulation.

Alternatively, assumed average heads, $H_t^o$, and flows, $Q_t^o$, can be used in a linear approximation of the non-separable product terms, $Q_t^T H_t$:

$$Q_t^T H_t = H_t^o Q_t^T + Q_t^o H_t - Q_t^o H_t^o \qquad (11.72)$$

Again, the model may need to be solved several times in order to identify reasonably accurate average flow and head estimates in each period $t$.

The amount of electrical energy produced is limited by the installed kilowatts of plant capacity, $P$, as well as on the plant factor, $p_t$. The plant factor is a measure of hydroelectric power plant use. Its value depends on the characteristics of the power system and the demand pattern for hydroelectric energy. The plant factor is defined as the average power load on the plant for the period divided by the installed plant capacity. The plant factor accounts for the variability in the demand for hydropower during each period $t$. This factor is usually specified by those responsible for energy production and distribution. It may or may not vary for different time periods.

The total energy produced cannot exceed the product of the plant factor, $p_t$, the number of hours, $h_t$, in the period, and the plant capacity, $P$, measured in kilowatts:

$$KWH_t \leq P h_t p_t \qquad (11.73)$$

## 3.5. Flood Risk Reduction

Two types of structural alternatives exist for flood risk reduction. One is flood storage capacity in reservoirs designed to reduce downstream peak flood flows. The other is channel enhancement and/or flood-proofing works that will contain peak flood flows and reduce damage. This section introduces methods of modelling both of these alternatives for inclusion in either benefit–cost or cost-effectiveness analyses. The latter analyses apply to situations in which a significant portion of the flood control benefits cannot be expressed in monetary terms and the aim is to provide a specified level of flood protection at minimum cost.

The discussion will first focus on the estimation of flood storage capacity in a single reservoir upstream of a potential flood damage site. This analysis will then be expanded to include downstream channel capacity improvements. Each of the modelling methods discussed will be appropriate for inclusion in multipurpose river basin planning (optimization) models having longer time step durations than those required to predict flood peak flows. Additional detail can be found in Appendix D.

### 3.5.1. Reservoir Flood Storage Capacity

A common approach for estimating reservoir flood storage capacity is based on expected damage reduction associated with various storage capacities in a particular reservoir. A relationship between flood storage capacity and expected flood damage reduction can be obtained. This functional relationship can then be used within a multipurpose reservoir model that considers other uses of the reservoir and its stored water. This approach avoids having to include short time steps applicable to most floods in an overall multipurpose river basin model.

Consider a reservoir upstream of a potential flood damage site along a river. The question is how much flood storage capacity, if any, should it contain. For various assumed capacities, simulation models can be used to predict the impact on the downstream flood peaks. These hydraulic simulation models must include flood routing procedures from the reservoir to the downstream potential damage site and the flood control operating policy at the reservoir. For various downstream flood peaks, economic flood damages can be estimated. To calculate the expected annual damages associated with any upstream reservoir capacity, the probabilities of various damage levels being exceeded in any year need to be calculated.

The expected annual flood damage at a potential flood damage site can be estimated from an exceedance probability distribution of peak flood flows at that potential damage site together with a flow or stage damage function. The peak flow exceedance distribution at any potential damage site will be a function of the upstream reservoir flood storage capacity $K_f$ and the reservoir operating policy. The computational process of determining the probability of exceedance of particular flood damages associated with a particular flood storage capacity and operating policy is illustrated graphically in Figure 11.22. The analysis requires three input functions that are shown in quadrants (a), (b) and (c). The dashed-line rectangles define point values on the three input functions in quadrants (a), (b) and (c) and the corresponding probabilities of exceeding a given level of damages in the lower right quadrant (d). The distribution in quadrant (d) is defined by the intersections of these dashed-line rectangles. This distribution defines the probability of equalling or exceeding a specified damage in any given year. The shaded area under the derived function is the annual expected damage, E[$FD$].

The relationships between flood stage and damage, and flood stage and peak flow, defined in quadrants (a) and (b) of Figure 11.22, must be known. These do not depend on the flood storage capacity in an upstream reservoir. The information in quadrant (c) defines the exceedance probabilities of each peak flow. Unlike the other two functions, this distribution depends on the upstream flood storage capacity and flood flow release policy. This peak flow probability of exceedance distribution is determined by simulating the annual floods entering the upstream reservoir in the years of record.

The difference between the expected annual flood damage without any upstream flood storage capacity and the expected annual flood damage associated with a flood storage capacity of $K_f$ is the expected annual flood damage reduction. This is illustrated in Figure 11.23. Knowing the expected annual damage reduction associated with various flood storage capacities, $K_f$, permits the definition of a flood damage reduction function, $B_f(K_f)$.

If the reservoir is a single-purpose flood control reservoir, the eventual tradeoff is between the expected flood reduction benefits, $B_f(K_f)$, and the annual costs, $C(K_f)$, of that upstream reservoir capacity. The particular reservoir flood storage capacity that maximizes the net benefits, $B_f(K_f) - C(K_f)$, may be appropriate from a national economic efficiency perspective but it may not be best from a local perspective. Those occupying the potential damage site may prefer a specified level of protection from that

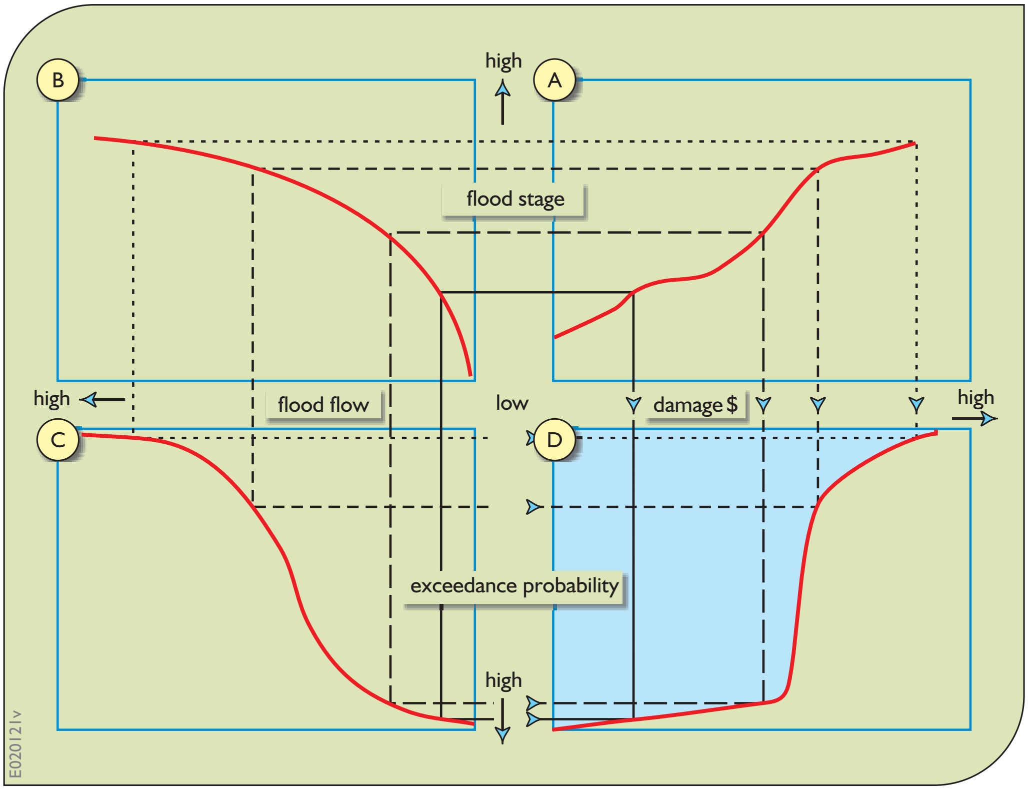



**Figure 11.22.** Calculation of the expected annual flood damage shown as the shaded area in quadrant (D) derived from the expected stage-damage function (A), the expected stage-flow relation (B), and the expected probability of exceeding an annual peak flow (C).

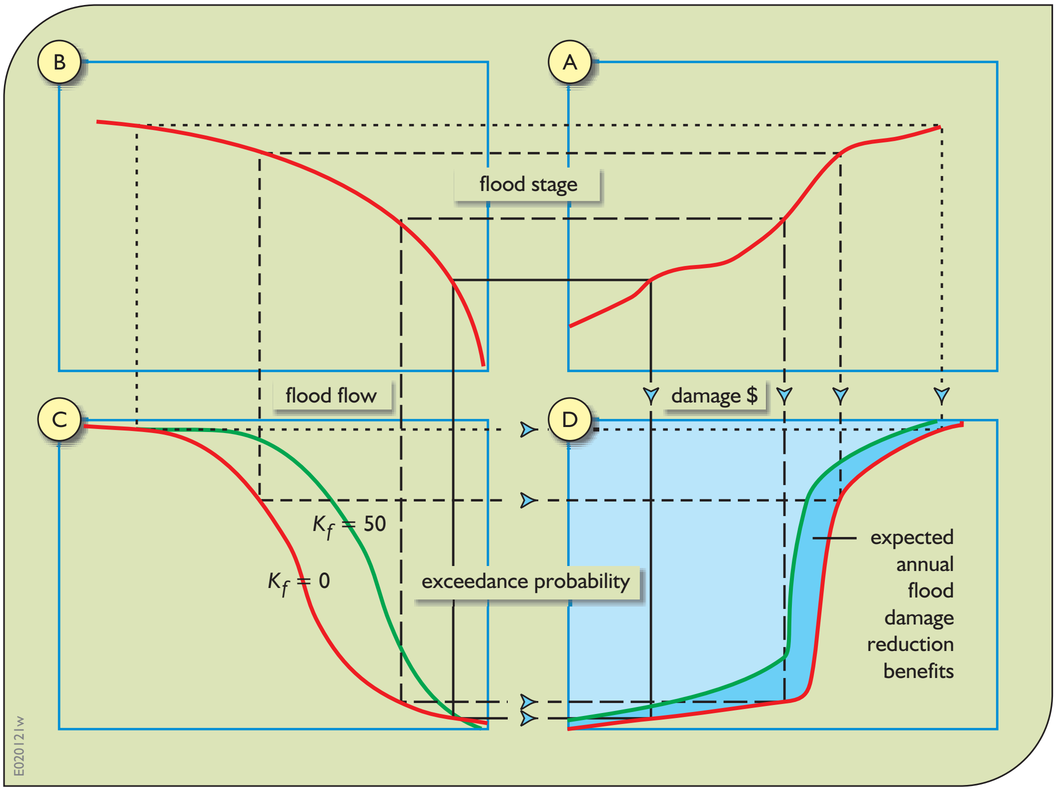



**Figure 11.23.** Calculation of expected annual flood damage reduction benefits, shown as the darkened portion of quadrant (D), associated with a specified reservoir flood storage capacity.

reservoir storage capacity, rather than the protection that maximizes expected annual net benefits, $B_f(K_f) - C(K_f)$.

If the upstream reservoir is to serve multiple purposes, say for water supply, hydropower and recreation as well as for flood control, then the expected flood reduction benefit function just derived could be a component in the overall objective function for that reservoir. Alternatively, the flood storage capacity, $K_f$, can be based on a particular predefined level of flood protection, regardless of the damage reduction. Again, both methods are discussed in greater detail in Appendix D.

Total reservoir capacity $K$ will equal the sum of dead storage capacity $K_d$, active storage capacity $K_a$ and flood storage capacity $K_f$, assuming they are the same in each period $t$. In some cases they may vary over the year. If the required active storage capacity can occupy the flood storage zone when flood protection is not needed, then the total reservoir capacity $K$ will be the dead storage, $K_d$, plus the maximum of either the actual storage volume and flood storage capacity in the flood season, or the actual storage volume in non-flood season.

$K \geq K_d + S_t + K_f$ for all periods t in flood season plus the following period that represents the end of the flood season (11.74)

$K \geq K_d + S_t$ for all remaining periods $t$ (11.75)

In the above equations, the dead storage capacity, $K_d$, is a known variable. It is included in the capacity Equations 11.74 and 11.75, assuming that the active storage capacity is greater than zero. Clearly, if the active storage capacity were zero, there would be no need for dead storage.

### 3.5.2. Channel Capacity

The unregulated natural peak flow of a particular design flood at a potential flood damage site can be reduced by upstream reservoir flood storage capacity, or it can be contained within the channel at the potential damage site by levees and other channel-capacity improvements. Of course both options may be applicable at some sites. What is needed is a function defining the reduction in the peak stage of some design flood and the cost of providing additional channel capacity, $K_c$. This function can then be included in a model along with the upstream flood storage capacity, $K_f$, to determine the best combination of alternative flood control measures that will protect the potential damage site from a predetermined design flood.

For example, consider a single-purpose flood control reservoir and channel capacity enlargement, either by dredging or dykes, whichever is cheaper, as alternatives for protecting from a design flood, $Q_T$, having a return period of $T$. Defining the functional relationship between $K_f$ and peak flow reduction at the damage site, and defining $K_c$ as the peak flow reduction due to channel capacity enlargement, then the least cost combination can be estimated from solving the following optimization model:

$$\text{minimize } \text{Cost}_R(K_f) + \text{Cost}_C(K_c) \tag{11.76}$$

subject to

$$Q_T \leq f_T(K_f) + K_c \tag{11.77}$$

Equations 11.76 and 11.77 assume that a decision will be made to provide protection from a design flood $Q_T$. It is only a question of how to provide the required protection.

Solving Equations 11.76 and 11.77 for peak flows $Q_T$ of various design floods of return period $T$ will identify the risk–cost tradeoff. This tradeoff function might look like what is shown in Figure 11.24.

## 3.6. Lake-Based Recreation

Recreation benefits derived from natural lakes as well as reservoirs are usually dependent on their storage levels. If docks, boathouses, shelters and other recreational facilities are installed on the basis of some assumed (target)

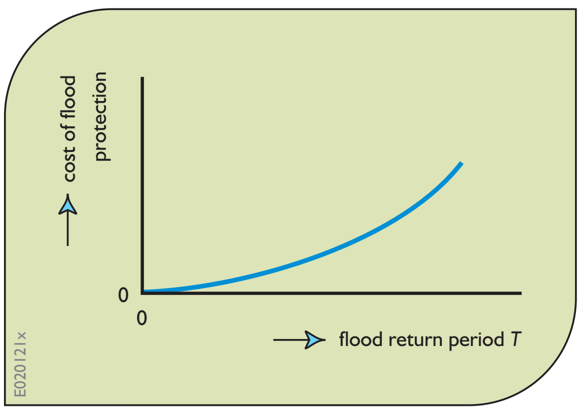



**Figure 11.24.** Tradeoff between minimum cost of flood protection and flood risk, as expressed by the expected return period.

lake level, and the lake levels deviate from the target value, there can be reduced recreational benefits. These storage targets and any deviations can be modelled in a way similar to Equation 11.66. The actual storage volume at the beginning of a recreation period $t$ equals the target storage volume less any deficit or plus any excess;

$$S_t^s = T^s - D_t^s + E_t^s \tag{11.78}$$

The recreational benefits in any recreational period $t$ can be defined on the basis of what they would be if the target were met, less the average of any losses that may occur from initial and final storage volume deviations, $D_t^s$ or $E_t^s$, from the target storage volume in each period of the recreation season (Chapter 10).

# 4. River Basin Analysis

## 4.1. Model Synthesis

Each of the model components discussed above can be combined, as applicable, into a model of a river system. One such river system together with some of its interested stakeholders is shown in Figure 11.25.

One of the first tasks in modelling this basin is to identify the actual and potential system components and their interdependencies. This is facilitated by drawing a schematic of the system at the level of detail that will address the issues being discussed and of concern to these stakeholders. This schematic can be drawn over the basin as in Figure 11.26. The schematic without the basin is illustrated in Figure 11.27.

A site number must be assigned at each point of interest. These sites are usually where some decision must be made. Mass-balance and other constraints will need to be defined at each of those sites.

**Figure 11.25.** A multipurpose river system whose management is of concern to numerous stakeholders (with permission from *Engineering News Record*, McGraw Hill).

**Figure 11.26.** A schematic representation of the basin components and their interdependencies drawn over the map image of the basin.

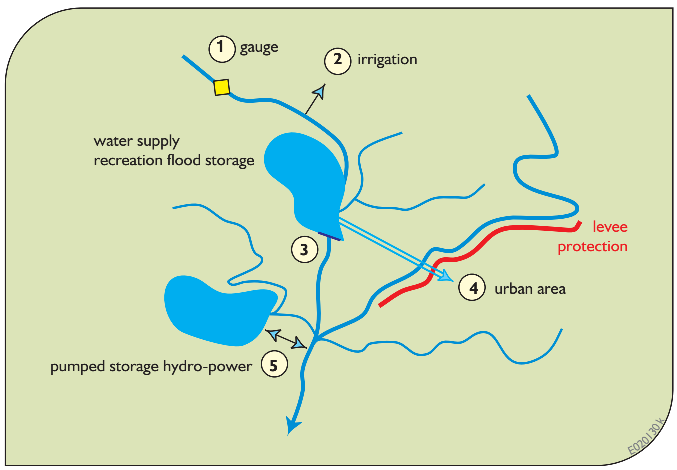



**Figure 11.27.** Schematic of river system showing components of interest at designated sites.

As shown in the schematic in Figure 11.27, this river has one streamflow gauge site, site 1, two reservoirs, sites 3 and 5, two diversions, sites 2 and 3, one hydropower plant, site 5, and a levee desired at site 4 to help protect against floods in the urban area. The reservoir at site 5 is a pumped storage facility. The upstream reservoir at site 3 is used for recreation, water supply, and flood control. The downstream reservoir is strictly for hydropower production.

Before developing a model of this river system, the number of within-year time periods $t$ to include in the model and the length of each within-year time period should be determined. If a river system's reservoirs are to contain storage for the distribution of water among different years, called over-year storage, then a number of periods encompassing multiple years of operation must be included in the model. This will allow an evaluation of the possible benefits of storing excess water in wet years for release in dry years.

Many reservoir systems completely fill up almost every year, and in such cases one is concerned only with the within-year operation of the system. This is the problem addressed here. To model the within-year operation of the system, a year is divided into a number of within-year periods. The number of periods and the duration of each period will depend on the variations in the hydrology, the demands and the particular objectives, as previously discussed.

## 4.2. Modelling Approach Using Optimzation

Once the number and duration of the time steps to be modelled have been identified, the variables and functions used at each site must be named. It is convenient to use notation that can be remembered when examining the model solutions. The notation made up for this example is shown in Table 11.5. Economic efficiency objective function components are defined in Table 11.6.

The objective components listed in Table 11.6 are economic efficiency objectives. The *LD* and *LE* loss functions are associated with deficit and excess allocations, respectively. These and the other functions (whose meaning should be self evident) need to be known, of course. There could, and no doubt should, be other objective components defined as well, as discussed in Chapter 10. Nevertheless these objective components serve the purpose here of illustrating how a model of this river system can be constructed.

| **design and operating variables and parameters** |
|---|
| **all sites *s* and time periods *t*:** |
| natural streamflows, based on gage flows at site 1, $Q^S(t)$ |
| **site 2** |
| irrigation allocations, $X^2(t)$, for all periods $t$ |
| annual irrigation target allocation, $T^2$ |
| known fraction of annual irrigation target required for each period $t$, $\delta_t^2$ |
| irrigation diversion channel capacity, $X^2$ |
| **site 3** |
| active initial storage volume. $S^3(t)$, in each period $t$ |
| dead storage volume, $K_d^3$ |
| reservoir release downstream, $R^3(t)$, in each period $t$ |
| recreation storage volume target, $T^3$ |
| deficit and excess storage volumes, $D^3(t)$ and $E^3(t)$, for each period $t$ |
| flood storage volume capacity, $K_f^3$ |
| total reservoir storage capacity, $K^3$ |
| urban diversions, $X^3(t)$, in each period $t$ |
| diversion capacity, $X^3$ |
| **site 4** |
| channel flood flow capacity, $Q^4$ |
| water supply target, $T^4$ |
| known fraction of annual water supply target for each period $t$, $\delta_t^4$ |
| **site 5** |
| active initial storage volume, $S^5(t)$, in each period $t$ |
| dead storage volume, $K_D^5$ |
| reservoir release through turbines, $QO^5(t)$, in each period $t$ |
| quantity of water pumped back into reservoir, $QI^5(t)$, in each period $t$ |
| energy produced, $EP^5(t)$, in each period $t$ |
| energy consumed, $EC^5(t)$, in each period $t$ |
| total storage capacity of reservoir, $K^5$ |
| power plant/pump capacity, $P^5$ |
| storage head function, $h(S^5(t))$, in each period $t$ |
| average storage head, $H^5(t)$, in each period $t$ |

E020912a

**Table 11.5. Names associated with required variables and functions at each site in Figure 11.27. The units of these variables and parameters, however defined, must be consistent.**

The overall objective might be a weighted combination of all objective components such as:

$$\text{Maximize} \sum_s w_s NB^s \tag{11.79}$$

This objective function does not identify how much each stakeholder group would benefit and how much they would pay. Who benefits and who pays, and how much, may be important. If it is known how much of each of the

| |
|---|
| **site 2** |
| $NB^2 = \text{IrrBenefit}^2(T^2) - \text{Cost}_D(X^2)$ |
| **site 3** |
| $NB^3 = \text{RecBenefit}^3(T^3) - \sum_t [LD^3(D^3(t)) + LE^3(E^3(t))] + FBenefit^3(K_f^3) - Cost_K{}^3(K^3)$ |
| **site 4** |
| $NB^4 = Div\text{Benefit}^4(T^4) - \sum_t [LD^4(D^4(t))] + FBenefit^4(Q^4) - Cost_Q(Q^4) - Cost_D(X^4)$ |
| **site 5** |
| $NB^5 = Energy\text{Benefit}^5(EP^5(t)) - EnergyCost^5(EC^5(t)) - Cost_p{}^5(P^5) - Cost_K{}^5(K^5)$ |

E020912b

**Table 11.6.** Economic efficiency (net benefit) maximization objective components.

net benefits derived from each site are to be allocated to each stakeholder group $i$, then these allocated portions, denoted as $NB_i^s$, of the total net benefits, $NB^s$, can be included in the overall objective:

$$\text{Maximize} \sum_i w_i \sum_s NB_i^s \tag{11.80}$$

Using methods discussed in Chapter 10, solving the model for various assumed values of these weights, $w_s$ or $w_i$, can help identify the tradeoffs between different conflicting objectives, Equation 11.79, or conflicting stakeholder interests, Equation 11.80.

The next step in model development is to define the constraints applicable at each site. It is convenient to begin at the most upstream sites and work downstream. As additional variables or functions are needed, invent notation for them. These constraints tie the decision-variables together and identify the interdependencies among system components. In this example, Equations 11.81 through 11.97 (see Box 11.2), together with objective Equation 11.79 or 11.80, define the general structure of this river system model.

Before the model can be solved the actual functions must be defined. Then they may have to be made piecewise linear if linear programming is to be the optimization procedure used to solve the model. The process of defining functions may add variables and constraints to the model, as discussed in Chapters 4 and 10.

For $T$ within-year periods $t$, this static model of a single year includes between $14T + 8$ and $16T + 5$ constraints, depending on the number of periods in the irrigation and recreation seasons. This number does not include the additional constraints that will surely be needed to define the functions in the objective function components and constraints. Models of this size and complexity, even though this is a rather simple river system, are usually solved using linear programming algorithms simply because other non-linear or dynamic programming (optimization) methods are more difficult to implement.

The model just developed is for a typical single year. In some cases it may be more appropriate to incorporate over-year as well as within-year mass-balance constraints, and yields with their respective reliabilities, within this modelling framework. This can be done as outlined in Sections 2.2.4 and 2.2.5 of this chapter.

The information derived from optimization models of river systems such as this one should not be considered as a final answer. Rather, it is an indication of system design and operation policies that should be further analysed using more detailed analyses. Optimization models of the type just developed serve as ways to eliminate inferior alternatives from further consideration rather than as ways of finding a solution that all stakeholders will accept as the best.

## 4.3. Modelling Approach Using Simulation

A simulation approach basically uses the same water balance equations and constraints as given in the text box.

**Box 11.2. General Structure of the River System Model**

**At site 1**

No constraints are needed at this gauge site.

**At site 2**

- The diverted water, $X^2(t)$, cannot exceed the streamflow, $Q^2(t)$, at that site.

$$Q^2(t) \geq X^2(t) \quad \forall t \text{ in the irrigation season} \tag{11.81}$$

- The diversion flow, $X^2(t)$, cannot exceed the diversion channel capacity, $X^2$.

$$X^2 \geq X^2(t) \quad \forall t \tag{11.82}$$

- The diversion flow, $X^2(t)$, must meet the irrigation target, $\delta_t^2 T^2$

$$X^2(t) \geq \delta_t^2 T^2 \quad \forall t \tag{11.83}$$

**At site 3**

- Storage volume mass balances (continuity of storage), assuming no losses.

$$S^3(t{+}1) = S^3(t) + Q^3(t) - X^3(t) - R^3(t) \quad \forall t, \; T + 1 = 1 \tag{11.84}$$

- Define storage deficits, $D^3(t)$, and excesses, $E^3(t)$, relative to recreation target, $T^3$.

$$S^3(t) = T^3 - D^3(t) + E^3(t) \quad \forall t \text{ in recreation season plus following period} \tag{11.85}$$

- Diverted water, $X^3(t)$, cannot exceed diversion channel capacity, $X^3$.

$$X^3(t) \leq X^3 \quad \forall t \tag{11.86}$$

- Reservoir storage capacity constraints involving dead storage, $K_D^3$, and flood storage, $K_F^3$, capacities.

$$S^3(t) \leq K^3 - K_D^3 - K_F^3 \quad \forall t \text{ in flood season plus following period}$$

$$S^3(t) \leq K^3 - K_D^3 \text{ for all other periods } t \tag{11.87}$$

**At site 4**

- Define deficit diversion, $D^4(t)$, from site 3, associated with target, $d_t^4 T^4$, if any.

$$X^3(t) = \delta_t^4 T^4 - D^4(t) \quad \forall t \tag{11.88}$$

- Channel capacity, $Q^4$, for peak flood flow, $PQ_T^4$, of specified return period $T$.

$$Q^4 \geq PQ_T^4 \tag{11.89}$$

**At site 5**

- Continuity of pumped storage volumes, involving inflows, $QI^5(t)$, and outflows, $QO^5(t)$, and assuming no losses.

$$S^5(t{+}1) = S^5(t) + QI^5(t) - QO^5(t) \quad \forall t \tag{11.90}$$

- Active storage capacity involving dead storage, $K_D^5$.

$$S^5(t) \leq K^5 - K_D^5 \quad \forall t \tag{11.91}$$

- Pumped inflows cannot exceed the amounts of water available at the intake; this includes the release from the upstream reservoir, $R^3(t)$, and the incremental flow, $Q^5(t) - Q^3(t)$.

$$QI^5(t) \leq Q^5(t) - Q^3(t) + R^3(t) \quad \forall t \tag{11.92}$$

- Define the energy produced, $EP^5(t)$, given the average storage head, $H(t)$, flow through the turbines, $QO^5(t)$, and efficiency, $e$.

$$EP^5(t) = (\text{const.})(H(t))(QO^5(t))e \quad \forall t \tag{11.93}$$

- Define the energy consumed, $EC^5(t)$, from pumping given the amount pumped, $QI^5(t)$.

$$EC^5(t) = (\text{const.})(H(t))(QI^5(t))/e \quad \forall t \tag{11.94}$$

- Energy production, $EP^5(t)$, and consumption, $EC^5(t)$, constraints given power plant capacity, $P^5$.

$$EP^5(t) \leq P^5 \text{ (hours of energy production in } t) \quad \forall t \tag{11.95}$$

$$EC^5(t) \leq P^5 \text{ (hours of pumping in } t) \quad \forall t \tag{11.96}$$

- Define the average storage head, $H^5(t)$, based on storage head functions, $h(S^5(t))$.

$$H^5(t) = (h(S^5(t{+}1) + h(S^5(t))/2 \quad \forall t \tag{11.97}$$

In a simulation approach, the analyst will have to specify the values of the decision variables (e.g. the diversion policies). The 'best' values of these decision-variables will have to be determined by sensitivity analyses, i.e. by changing these variables and seeing what the output will be in terms of the objective functions. Hence, simulation is a 'trial-and-error' approach.

A simulation model of the simple system as given in Figure 11.27 can easily be developed by means of a spreadsheet. More complicated systems can be simulated by using generic computer packages for river basin planning such as RIBASIM (WL | Delft Hydraulics, 2004), MIKE BASIN (DHI, 2003), and WEAP-21 (SEI, 2001). These river basin simulation packages support the development of a model schematization consisting of a network of nodes connected by links. The nodes represent reservoirs, dams, weirs, pumps, hydro-power stations, water users, inflows, artificial and natural bifurcations, intake structures, natural lakes and so on. The links transport water between the different nodes. Such a network represents the basin's features that are significant for the planning and management problem at issue. The network can be adjusted to provide the level of spatial and temporal detail required. The river basin is represented as a network schematization superimposed over a vector or raster map image.

Figure 11.28 shows the network schematization of RIBASIM of the simple river basin system of Figure 11.27. Within each time-step a water balance calculation is made in two phases:

1. The target setting phase, in which the water demands are determined (calculated from specific modules for irrigation, DMI, and other uses, or directly specified in volumes or discharges), resulting in targets for the releases from surface water reservoirs, aquifers and diversion flows at weirs and pumping stations.
2. The water allocation phase, in which the allocation to the users takes place according to the targets, availability of supply and allocation rules (priorities).

Water allocation to users can be implemented in a variety of ways. In its simplest format, water is allocated on a 'first come, first serve' principle along the natural flow direction. This allocation can be amended by rules which, for

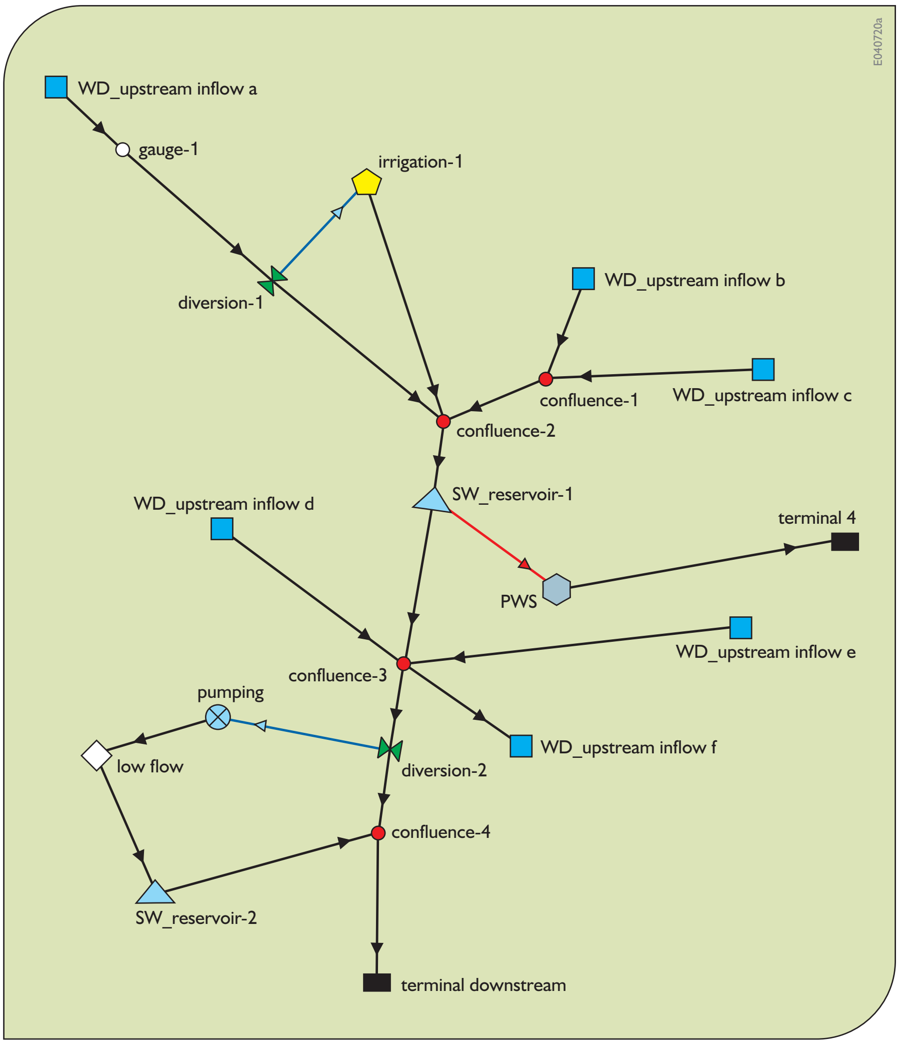



**Figure 11.28.** RIBASIM representation the river basin schematization of Figure 11.27.

example, allocate priority to particular users or result in allocations proportional to demands. On the basis of a set of simulations in which the values of important decision-variables are changed, thus defining a range of alternative development or management strategies, thusthe performance of the basin is evaluated in terms of water allocation, shortages, energy production, overall river basin water balance, crop yields, production costs and so on.

## 4.4. Optimization and/or Simulation

For determining the best design and operating policy variable values in river basin systems such as the one just illustrated, or even more complex ones, the use of both optimization and simulation models can be advantageous. As discussed in Chapter 3 (Section 2.2), optimization is often useful, not for finding the very best solution, but for eliminating the worst alternatives from further consideration. The remaining ones can then be analysed in greater detail, using more detailed simulation models. Simulation by itself begs the question: 'What to simulate?' Optimization by itself begs the question: 'Is the solution really the best?'

There is an art to using optimization as a preliminary screening tool. Some of this art is illustrated in Chapter 4, Section 5.3.

## 4.5. Project Scheduling

The river basin models discussed thus far in this chapter deal with static planning situations. Project capacities, targets and operating policies take on fixed values and one examines 'snapshot steady-state' solutions for a particular year in the future. These 'snapshots' only allow for fluctuations caused by hydrological variability. The non-hydrological world is seldom static, however. Demands and targets change in response to population growth, investment in agriculture and industry, and shifting priorities for water use. In addition, financial resources available for water resources investment are limited and may vary from year to year.

Dynamic planning models can aid those responsible for the long-run development and expansion of water resources systems. Although static models can identify target values and system configuration designs for a particular period in the future, they are not well adapted to long-run capacity expansion planning over a ten, twenty or thirty-year period. But static models may identify projects for implementation in early years that in later year simulations do not appear in the solutions. This was seen in some examples presented in Chapter 4.

This is a common problem in capacity expansion where each project has a fixed construction and implementation cost, as well as variable operating, repair and maintenance cost components. If there are two mutually exclusive competing projects, one may be preferred at a site when the demand at that site is low, but the other may be preferred if the demand is, as it is later projected to be, much higher. Which of the two projects should be selected now when the demand is low, given the uncertainty of the projected increase over time, especially assuming it makes no economic sense to destroy and replace a project already built?

Whereas static models consider how a water resources system operates under a single set of fixed conditions, dynamic expansion models must consider the sequence of changing conditions that might occur over the planning period. For this reason, dynamic expansion models are potentially more complex and larger than are their static counterparts. However, to keep the size and cost of dynamic models within the limitations of most studies, these models are generally restricted to very simple descriptions of the economic and hydrological variables of concern. Most models use deterministic hydrology and are constrained either to stay within predetermined investment budget constraints or to meet predetermined future demand estimates.

Dynamic expansion models can be viewed as network models for solution by linear or dynamic programming methods. The challenge in river system capacity expansion or project scheduling models is that each component's performance, or benefits, may depend on the design and operating characteristics of other components in the system. River basin project impacts tend to be dependent on what else is happening in the basin, that is, what other projects are present and how they are operated.

Consider a situation in which $n$ fixed-scale discrete projects may be built during the planning period. The scheduling problem is to determine which projects to build or implement and in what order. The solution of this problem generally requires a resolution of the timing problem. When should each project be built or implemented, if at all?

For example, assume there are $n = 3$ discrete projects that might be beneficial to implement sometime over the next twenty years. This twenty-year period consists of four five-year construction periods $y$. The actual benefits derived from any new project may depend on the projects that already exist. Let $\boldsymbol{S}$ be the set of projects existing at the beginning of any construction period. Finally, let $NB_y(\boldsymbol{S})$ be the maximum present value of the total net benefits derived in construction period $y$ associated with the projects in the set $\boldsymbol{S}$. Here, 'benefits' refers to any composite of system performance measures.

These benefit values for various combinations of discrete projects could be obtained from static river system models, solved for all combinations of discrete projects for conditions existing at the end of each of these four five-year periods. It might be possible to do this for just one or two of these four periods and apply applicable discount rates for the other periods. These static models can be similar to those discussed in the previous section of this chapter. Now the challenge is to find the sequencing of these three projects over the periods $y$ that meet budget constraints and that maximize the total present value of benefits.

This problem can be visualized as a network. As shown in Figure 11.29, the nodes of this network represent the sets $\boldsymbol{S}$ of projects that exist at the beginning of the construction period. For these sets $\boldsymbol{S}$, we have the present value of their benefits, $NB_y(\boldsymbol{S})$, in the next five years. The links represent the project or projects implemented in that construction period. Any set of new projects that exceeds the construction funds available for that period is not shown on the network. Those links are infeasible. For the purposes of this example, assume it is not financially feasible to add more than one project in any single construction period. Let $C_{ky}$ be the present value of the cost of implementing project $k$ in construction period $y$.

The problem is to find the best (maximum benefits less costs) paths through the network. Each link represents a net benefit, $NB_y(\boldsymbol{S})$, over the next five-years obtained from the set of projects, $\boldsymbol{S}$, that exist less the cost of adding a new project $k$.

Using linear programming, define a continuous non-negative unknown decision-variable $X_{ij}$ for each link between node $i$ and node $j$. It will be an indicator of whether a link is on the optimum path or not. If after solving the model its value is 1, then the link connecting nodes $i$ and $j$ represents the decision to make in that construction period. Otherwise, its value is 0 indicating the link is not on the optimal path. The sequence of links having their $X_{ij}$ values equal to 1 will indicate the most beneficial sequence of project implementations.

Let the net benefits associated with node $i$ be designated $NB_i$ (which equals the appropriate $NB_y(\boldsymbol{S})$ value), and $C_{ij}$ the cost, $C_{ky}$, of the new project $k$ associated with that link. The objective is to maximize the present value of net benefits less project implementation costs over all periods $y$.

$$\text{Maximise} \sum_i \sum_j (NB_i - C_{ij}) X_{ij} \tag{11.98}$$

Subject to:

Continuity at each node:

$$\sum_h X_{hi} = \sum_j X_{ij} \text{ for each node } i \text{ in the network.} \tag{11.99}$$

Sum of all decision-variable values on the links in any one period $y$ must be 1. For example, in period 1:

$$X_{00}+X_{01}+X_{02}+X_{03}=1 \tag{11.100}$$

The sums in Equation 11.99 are over nodes $h$ having links to node $i$ and over nodes $j$ having links from node $i$.

The optimal path through this network can also be solved using dynamic programming. (Refer to the capacity expansion problem illustrated in Chapter 4). For a backward-moving solution procedure, let

$\boldsymbol{s}$ = subset of projects $k$ not contained in the set $\boldsymbol{S}$ ($\boldsymbol{s} \notin \boldsymbol{S}$).

$\$_y$ = the maximum project implementation funds available in period y.

$F_y(\boldsymbol{S})$ = the present value of the total benefits over the remaining periods, $y, y+1, \ldots, 4$.

$F_Y(\boldsymbol{S}) = 0$ for all sets of projects $S$ following the end of the last period.

The recursive equations for each construction period, beginning with the last period, can be written

$$F_y(\boldsymbol{S}) = \underset{\boldsymbol{s} \in \boldsymbol{S}}{\text{maximum}} \{NB_y(\boldsymbol{S}) - \sum_{k \in \boldsymbol{s}} C_{ky} + F_{y+1}(\boldsymbol{S} + \boldsymbol{s})\} \;\forall \boldsymbol{S} \tag{11.101}$$

$$\sum_{k \in \boldsymbol{s}} C_{ky} \leq \$_y$$

Defining $F'_y(\boldsymbol{S})$ as the present value of the total benefits of all new projects in the set $\boldsymbol{S}$ implemented in all periods up

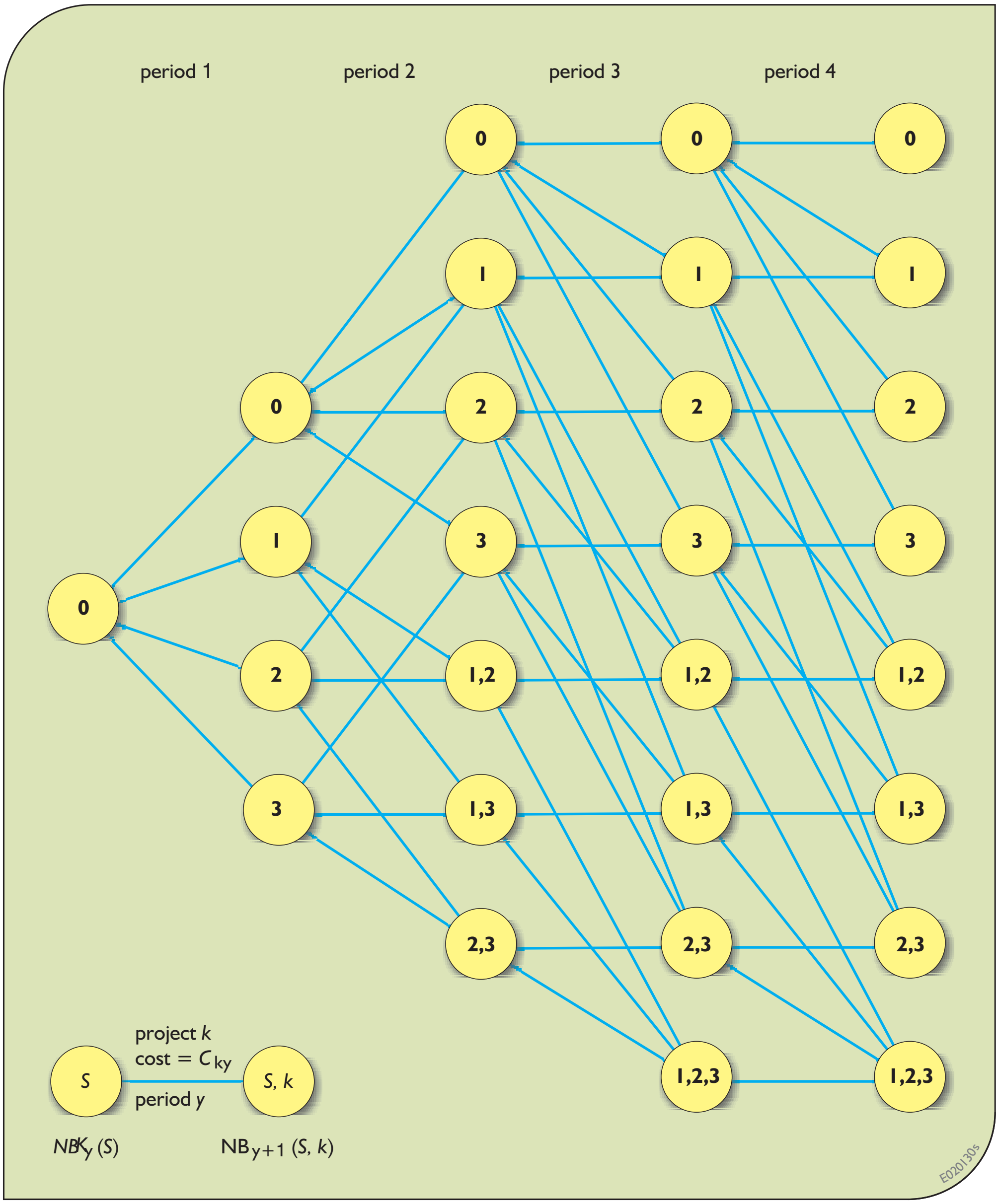



**Figure 11.29.** Project scheduling options. Numbers in nodes represent existing projects. Links represent new projects, the difference between the existing projects at both connecting nodes.

to and including period $y$, and the subset $s$ of projects $k$ being considered in period $y$ now belonging to the set $\boldsymbol{S}$ of projects existing at the end of the period, the recursive equations for a forward-moving solution procedure beginning with the first period, can be written:

$$F_y(\boldsymbol{S}) = \underset{\boldsymbol{s} \in \boldsymbol{S}}{\text{maximum}} \{NB_y(\boldsymbol{S}) - \sum_{k \in \boldsymbol{s}} C_{ky} + F_{y+1}(\boldsymbol{S} + \boldsymbol{s})\} \ \forall \boldsymbol{S} \tag{11.102}$$

$$\sum_{k \in \boldsymbol{s}} C_{ky} \leq \$_y$$

where $F_0'(0) = 0$.

Like the linear programming model, the solutions of these dynamic programming models identify the sequencing of projects recognizing their interdependencies. Of interest, again, is what to do in this first construction period. The only reason for looking into the future is to make sure, as best as one can, that the first period's decisions are not myopic. Models like these can be developed and solved again with more updated estimates of future conditions when next needed.

Additional constraints and variables might be added to these scheduling models to enforce requirements that some projects precede others, or that if one project is built another is infeasible. These additional restrictions usually reduce the size of a network of feasible nodes and links, as shown in Figure 11.29.

Another issue that these dynamic models can address is the sizing or capacity expansion problem. Frequently, the scale or capacity of a reservoir, pipeline, pumping station or irrigation project is variable and needs to be determined concurrently with the solution of the scheduling and timing problems. To solve the sizing problem, the costs and capacities in the scheduling model become variables.

## 5. Conclusions

This chapter on river basin planning models has introduced some ways of modelling river basin components, separately and together within an integrated model. Ignored during the development of these different model types were the uncertainties associated with the results of these models. As discussed in Chapters 7 through 9, these uncertainties may have a substantial effect on the model solution and the decision taken.

Most of this chapter has been focused on the development of simplified screening models, using simulation as well as optimization methods, for identifying what and where and when infrastructure projects should be implemented, and of what capacity. The solution of these screening models, and any associated sensitivity and uncertainty analyses, should be of value prior to committing to a more costly design modelling exercise.

Preliminary screening of river basin systems, especially given multiple objectives, is a challenge to accomplish in an efficient and effective manner. The modelling methods and approaches discussed in this chapter serve as an introduction to that art.

## 6. References

DHI (Danish Hydraulic Institute). 2003. *MIKE BASIN: a toolfor river basin planning and management. User Manual for MIKE BASIN 2003.* Horsholm, Denmark, Danish Hydraulic Institute. Also http://www.dhisoftware.com/mikebasin/index.htm (accessed 11 November 2004).

FAO. 1998. *Crop evapotranspiration: guidelines for computing crop water requirements*. R.G. Allen, L.S. Pereira, D. Raes and M. Smith. FAO Irrigation and Drainage Paper 56. Rome, FAO.

MAIDMENT, D.R. 1993. *Handbook of hydrology*. New York, McGraw-Hill.

RIPPL, W. 1883. The capacity of storage reservoirs for water supply. *Proceedings of the Institute of Civil Engineers* (UK), Vol. 71, pp. 270–8.

SEI, WEAP. 2001. *User guide for WEAP21*. Boston Mass., Stockholm Environment Institute, Tellus Institute, July; also http://www.weap21.org/index.sp (accessed 11 November 2004).

USDA (US Department of Agriculture). 1972. *National engineering handbook*. Washington, D.C., Soil Conservation Service, US Government Printing Office.

USDOE (US Departement of Energy). 2002. *Hydropower program.* http://hydropower.inel.gov or http://hydropower.id.doe. gov/, accessed April.

WCD (World Commission on Dams). 2000. *Dams and development: a new framework for decision-making*. The report of the World Commission on Dams. London, Earthscan.

WL | DELFT HYDRAULICS. 2004. *RIBASIM Version 6.32: Technical Reference Manual*. Also http://www.wldelft.nl/soft/ribasim/int/index.html (accessed 11 November 2004).

### Additional References (Further Reading)

BASSON, M.S.; ALLEN, R.B.; PEGRAM, G.G.S. and VAN ROOYEN, J.A. 1994. *Probabilistic management of water resource and hydropower systems*. Highlands Ranch, Colo., Water Resources Publications.

BECKER, L. and YEH, W.W.G. 1974. Optimal timing, sequencing, and sizing of multiple reservoir surface water supply facilities. *Water Resources Research*, Vol. 10, No. 1, pp. 57–62.

BECKER, L. and YEH, W.W.G. 1974. Timing and sizing of complex water resource systems. *Journal of the Hydraulics Division, ASCE*, Vol. 100, No. HYIO,pp. 1457–70.

BEDIENT, P.B. and HUBER, W.C. 1992. *Hydrology and floodplain analysis*, 2nd edn. Reading, Mass., Addison Wesley.

CHIN, D.A. 2000. *Water-resources engineering*. Upper Saddle River, N.J., Prentice Hall.

ERLENKOTTER, D. 1973a. Sequencing expansion projects. *Operations Research*, Vol. 21, No. 2, pp. 542–53.

ERLENKOTTER, D. 1973b. Sequencing of interdependent hydroelectric projects. *Water Resources Research*, Vol. 9, No. 1, pp. 21–7.

ERLENKOTTER, D. 1976. Coordinating scale and sequencing decisions for water resources projects: economic modelling for water policy evaluation. *NorthHolland/TIMS Studies in the Management Sciences*, Vol. 3, pp. 97–112.

ERLENKOTTER, D. and ROGERS, J.S. 1977. Sequencing competitive expansion projects. *Operations Research*, Vol. 25, No. 6, pp. 937–51.

HALL, W.A.; TAUXE, G.W. and YEH, W.W.G. 1969. An alternative procedure for optimization of operations for planning with multiple river, multiple purpose systems. *Water Resources Research*, Vol. 5, No. 6, pp. 1367–72.

JACOBY, H.D. and LOUCKS, D.P. 1972. Combined use of optimization and simulation models in river basin planning. *Water Resources Research*, Vol. 8, No. 6, pp. 1401–14.

JAMES, L.D. and LEE, R.R. 1971. Economics of water resources planning. New York, McGraw-Hill.

LINSLEY, R.K.; FRANZINI, J.B.; FREYBERG, D.L. and TCHOBANOGLOUS, G. 1992. Water-resources engineering. New York, McGraw-Hill.

LOUCKS, D.P. 1976. Surface water quantity management, In: A. K. Biswas (ed.), *Systems approach to water management, Chapter 5*. New York, McGraw-Hill.

LOUCKS, D.P.; STEDINGER, J.R. and HAITH, D.A. 1981. *Water resources systems planning and analysis*. Englewood Cliffs, N.J., Prentice Hall.

MAASS, A.; HUFSCHMIDT, M.M.; DORFMAN, R.; THOMAS, H.A. Jr.; MARGLIN, S.A. and FAIR, G.M. 1962. *Design of water resource systems*. Cambridge, Mass., Harvard University Press.

MAJOR, D.C. and LENTON, R.L. 1979. *Applied water resources systems planning. Englewood Cliffs*, N.J., Prentice Hall.

MAYS, L.W. 2005. *Water resources engineering*. New York, Wiley.

MAYS, L.W. and TUNG, Y.K. 1992. *Hydrosystems engineering and management*. New York, McGraw-Hill.

MORIN, T.L. 1973. Optimal sequencing of capacity expansion projects. *Journal of the Hydraulics Division, ASCE*, Vol. 99, No. HY9, pp. 1605–22.

MORIN, T.L. and ESOGBUE, A.M.O. 1974. A useful theorem in the dynamic programming solution of sequencing and scheduling problems occurring in capital expenditure planning. *Water Resources Research*, Vol. 10, No. 1, pp. 49–50.

NATIONAL RESEARCH COUNCIL. 1999. New strategies for *America's watersheds*. Washington, D.C., National Academy Press.

NATIONAL RESEARCH COUNCIL. 2000. *Risk analysis and uncertainty in flood damage reduction studies*. Washington, D.C., National Academy Press.

O'LAOGHAIRE, D.T. and HIMMELBLAU, D.M. 1974. *Optimal expansion of a water resources system*. New York, Academic Press.

REVELLE, C. 1999. *Optimizing reservoir resources*. New York, Wiley.

THOMAS, H.A. Jr. and BURDEN, R.P. 1963. *Operations research in water quality management*. Cambridge, Mass., Harvard Water Resources Group, 1963.

USACE. 1991a. *Hydrology and hydraulics workshop on riverine levee freeboard*. Monticello, Minnesota, Report SP-24. Hydrological Engineering Center, US Army Corps of Engineers.

USACE. 1991b. *Benefit determination involving existing levees*. Policy Guidance Letter No. 26. Washington, D.C., Headquarters, US Army Corps of Engineers.

USACE. 1999. *Risk analysis in geotechnical engineering for support of planning studies*. ETL 1110-2-556. Washington, D.C., Headquarters, US Army Corps of Engineers.

VIESSMAN, W. Jr. and WELTY, C. 1985. *Water management technology and institutions*. New York, Harper and Row.

WURBS, R.A. and JAMES, W.P. 2002. *Water resources engineering*. Upper Saddle River, N.J., Prentice Hall.

# 12. Water Quality Modelling and Prediction

1. Introduction 377

2. Establishing Ambient Water Quality Standards 378
   2.1. Water-Use Criteria 379

3. Water Quality Model Use 379
   3.1. Model Selection Criteria 380
   3.2. Model Chains 381
   3.3. Model Data 382

4. Water Quality Model Processes 383
   4.1. Mass-Balance Principles 384
      4.1.1. Advective Transport 385
      4.1.2. Dispersive Transport 385
      4.1.3. Mass Transport by Advection and Dispersion 385
   4.2. Steady-State Models 386
   4.3. Design Streamflows for Water Quality 388
   4.4. Temperature 389
   4.5. Sources and Sinks 390
   4.6. First-Order Constituents 390
   4.7. Dissolved Oxygen 390
   4.8. Nutrients and Eutrophication 393
   4.9. Toxic Chemicals 396
      4.9.1. Adsorbed and Dissolved Pollutants 396
      4.9.2. Heavy Metals 398
      4.9.3. Organic Micro-pollutants 399
      4.9.4. Radioactive Substances 400
   4.10. Sediments 400
      4.10.1. Processes and Modelling Assumptions 401
      4.10.2. Sedimentation 401
      4.10.3. Resuspension 401
      4.10.4. Burial 402
      4.10.5. Bed Shear Stress 402
   4.11. Lakes and Reservoirs 403
      4.11.1. Downstream Characteristics 405
      4.11.2. Lake Quality Models 406
      4.11.3. Stratified Impoundments 407

5. An Algal Biomass Prediction Model 408
   5.1. Nutrient Cycling 408
   5.2. Mineralization of Detritus 408
   5.3. Settling of Detritus and Inorganic Particulate Phosphorus 409
   5.4. Resuspension of Detritus and Inorganic Particulate Phosphorus 409
   5.5. The Nitrogen Cycle 409
      5.5.1. Nitrification and Denitrification 409
      5.5.2. Inorganic Nitrogen 410

5.6. Phosphorus Cycle 410
5.7. Silica Cycle 411
5.8. Summary of Nutrient Cycles 411
5.9. Algae Modelling 412
5.9.1. Algae Species Concentrations 412
5.9.2. Nutrient Recycling 413
5.9.3. Energy Limitation 413
5.9.4. Growth Limits 414
5.9.5. Mortality Limits 414
5.9.6. Oxygen-Related Processes 415

6. Simulation Methods 416
6.1. Numerical Accuracy 416
6.2. Traditional Approach 417
6.3. Backtracking Approach 418
6.4. Model Uncertainty 420

7. Conclusions: Implementing a Water Quality Management Policy 421

8. References 422

# 12 Water Quality Modelling and Prediction

The most fundamental human needs for water are for drinking, cooking and personal hygiene. To meet these needs, the quality of the water used must pose no risk to human health. The quality of the water in nature also affects the condition of ecosystems that all living organisms depend on. At the same time, humans use water bodies as convenient sinks for the disposal of domestic, industrial and agricultural wastewaters. This of course degrades the quality of those water bodies. Water resources management involves the monitoring and management of water quality as much as the monitoring and management of water quantity. Various models have been developed to assist in predicting the water quality impacts of alternative land and water management policies and practices. This chapter introduces some of the main principles of water quality modelling.

## 1. Introduction

Water quality management is a critical component of overall integrated water resources management. Most users of water depend on adequate levels of water quality. When these levels are not met, these water users must either pay an additional cost for water treatment or incur at least increased risks of damage or loss. As populations and economies grow, more pollutants are generated. Many of these are waterborne, and hence can end up in surface and groundwater bodies. Increasingly, the major efforts and costs involved in water management are devoted to water quality protection and management. Conflicts among various users of water are increasingly over issues involving water quality as well as water quantity.

Natural water bodies are able to serve many uses, including the transport and assimilation of waterborne wastes. But as natural water bodies assimilate these wastes, their quality changes. If the quality drops to the extent that other beneficial uses are adversely affected, the assimilative capacities of those water bodies have been exceeded with respect to those affected uses. Water quality management measures are actions taken to ensure that the total pollutant loads discharged into receiving water bodies do not exceed the ability of those water bodies to assimilate those loads while maintaining the levels of quality specified by quality standards set for those waters.

What uses depend on water quality? One can identify almost any use. All living organisms require water of sufficient quantity and quality to survive, although different aquatic species can tolerate different levels of water quality. Regrettably, in most parts of the developed world it is no longer safe to drink natural surface or ground waters; they usually need to be treated before they become fit for human consumption. Treatment is not a practical option for recreational bathing, or for maintaining the health of fish, shellfish and other organisms found in natural aquatic ecosystems. Thus, standards specifying minimum acceptable levels of quality are set for most ambient waters. Various other uses have their own standards as well. Irrigation water must not be too saline or contain toxic substances that can be absorbed by the plants or destroy microorganisms in the soil. Water quality standards for industry can be very demanding, depending of course on the particular industrial processes.

Pollutant loadings degrade water quality. High domestic wasteloads can result in high concentrations of bacteria, viruses and other organisms that affect human health. High organic loadings can reduce dissolved oxygen to levels that are fatal to parts of the aquatic ecosystem and cause obnoxious odours. Nutrient loadings from both urban and agricultural land runoff can cause excessive algae growth, which in turn may degrade the water aesthetically and recreationally, and upon death result in low dissolved oxygen levels. Toxic heavy metals and other micro-pollutants can accumulate in the bodies of aquatic organisms, including fish, making them unfit for human consumption even if they themselves survive.

Pollutant discharges originate from point and non-point sources. A common approach to controlling point source discharges, such as those from stormwater outfalls, municipal wastewater treatment plants or industries, is to impose standards specifying maximum allowable pollutant loads or concentrations in their effluents. This is often done in ways that are not economically efficient or even environmentally effective. Effluent standards typically do not take into account the particular assimilative capacities of the receiving water body.

Non-point sources such as agricultural runoff or atmospheric deposition are less easily controlled, and hence it is difficult to apply effluent standards to non-point source pollutants, and their loadings can be much more significant than point source loadings. Management of non-point water quality impacts requires a more ambient-focused water quality management programme.

The goal of an ambient water quality management programme is to establish appropriate standards for water quality in water bodies receiving pollutant loads, and then to ensure that these standards are met. Realistic standard-setting takes into account the basin's hydrological, ecological and land use conditions, the potential uses of the receiving water body, and the institutional capacity to set and enforce water quality standards.

Ambient-based water quality prediction and management involves considerable uncertainty. No one can predict what pollutant loadings will occur in the future, especially from area-wide non-point sources. In addition to uncertainties inherent in measuring the attainment of water quality standards, there are uncertainties in models used to determine sources of pollution, to allocate pollutant loads, and to predict the effectiveness of actions taken to meet water quality standards. The models available to help managers predict water quality impacts are relatively simple compared with the complexities of actual water systems. These limitations and uncertainties should be understood and addressed as water quality management decisions are based on their outputs.

## 2. Establishing Ambient Water Quality Standards

Identifying the intended uses of a water body – whether a lake, a section of a stream or an estuary – is a first step in setting water quality standards for that body. The most restrictive of the specific desired uses of a water body is termed a *designated use*. Barriers to achieving the designated use are the presence of pollutants, or hydrological and geomorphic changes that affect the water quality.

The designated use dictates the appropriate type of water quality standard. For example, a designated use of human recreation should protect humans from exposure to microbial pathogens while swimming, wading or boating. Other uses include those designed to protect humans and wildlife from consuming harmful substances in water, in fish and in shellfish. Aquatic-life uses include the protection and propagation of fish, shellfish and wildlife resources.

Standards set upstream may affect the uses of water downstream. For example, small headwater streams may have aesthetic value but may not be able to support extensive recreational uses. However, their condition may affect the ability of a downstream area to achieve a particular designated use such as 'fishable' or 'swimmable'. In this case, the designated use for the smaller upstream water body may be defined in terms of achieving the designated use of the larger downstream water body.

In many areas, human activities have altered the landscape and aquatic ecosystems to the point where they cannot be restored to their pre-disturbance condition. For example, someone's desire to establish a trout fish-farm in downtown Paris, Phnom Penh, Prague or Pretoria may not be attainable because of the development history of these areas or the altered hydrological regimes of the rivers flowing through them. Similarly, someone might wish to designate an area near the outfall of a sewage treatment plant for shellfish harvesting, but health considerations would preclude any such use. Ambient water quality standards must be realistic.

Designating the appropriate use for a water body is a policy decision that can be informed by the use of water quality prediction models of the type discussed in this chapter. However, the final standard selection should reflect a social consensus made while bearing in mind the current condition of the watershed, its pre-disturbance condition, the advantages derived from a certain designated use, and the costs of achieving that use.

### 2.1. Water-Use Criteria

The designated use is a qualitative description of the desired condition of a water body. A criterion is a measurable indicator surrogate for use attainment. The criterion may be positioned at any point in the causal chain of boxes shown in Figure 12.1.

In Box 1 of Figure 12.1 are measures of the pollutant discharge from a treatment plant (such as biological oxygen demand, ammonia ($NH_3$), pathogens and suspended sediments) or the amount of a pollutant entering the edge of a stream from runoff. A criterion at this position is referred to as an effluent standard. Criteria in Boxes 2 and 3 are possible measures of ambient water quality conditions. Box 2 includes measures of a water quality parameter such as dissolved oxygen (*DO*), pH, total phosphorus concentration, suspended sediment or temperature. Criteria closer to the designated use (e.g. Box 3)

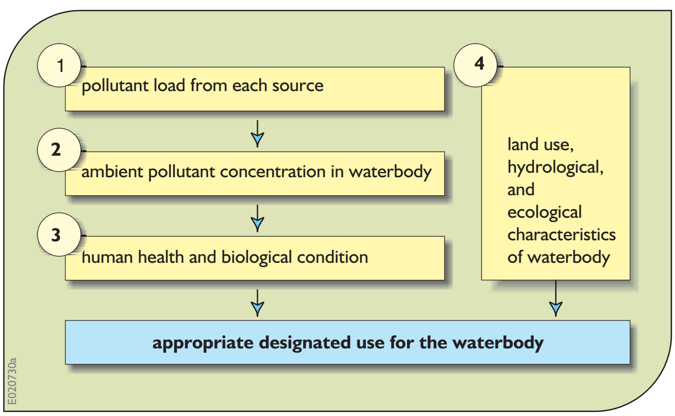



**Figure 12.1.** Factors considered when determining designated use and associated water quality standards.

include more combined or comprehensive measures of the biological community as a whole, such as the condition of the algal community (chlorophyll *a*) or a measure of contaminant concentration in fish tissue. Box 4 represents criteria that are associated with sources of pollution other than pollutants. These criteria might include measures such as flow timing and pattern (a hydrological criterion), abundance of non-indigenous taxa, or some quantification of channel modification (e.g. a decrease in sinuosity) (NRC, 2001).

The more precise the statement of the designated use, the more accurate the criterion will be as an indicator of that use. For example, the criterion of fecal coliform count may be a suitable criterion for water contact recreation. The maximum allowable count itself may differ among water bodies that have water contact as their designated use, however.

Surrogate indicators are often selected for use as criteria because they are easy to measure and in some cases are politically appealing. Although a surrogate indicator may have these appealing attributes, its usefulness can be limited unless it can be logically related to a designated use.

As with setting designated uses, the connections among water bodies and segments must be considered when determining criteria. For example, where a segment of a water body is designated as a mixing zone for a pollutant discharge, the criterion adopted should assure that the mixing zone use will not adversely affect the surrounding water body uses. Similarly, the desired condition of a small headwater stream may need to be specified in relation to other water bodies downstream; thus, an ambient nutrient criterion may be set in a small headwater stream to ensure a designated use in a downstream estuary, even if there are no local adverse impacts resulting from the nutrients in the small headwater stream, as previously discussed. Conversely, a high fecal coliform criterion may be permitted upstream of a recreational area if the fecal load dissipates before the flow reaches that area.

## 3. Water Quality Model Use

Monitoring data are the preferred form of information for identifying impaired waters (Appendix B). Model predictions might be used in addition to or instead of monitoring data for several reasons:

- Modelling might be feasible in some situations where monitoring is not.
- Integrated monitoring and modelling systems could provide better information than one or the other alone for the same total cost. For example, regression analyses that correlate pollutant concentration with some more easily measurable factor (such as streamflow) could be used to extend monitoring data for preliminary listing (of impared status) purposes. Models can also be used in a Bayesian framework to determine preliminary probability distributions of impairment that can help direct monitoring efforts and reduce the quantity of monitoring data needed for making listing decisions at a given level of reliability (see Chapter 7).
- Modelling can be used to assess (predict) future water quality situations resulting from different management strategies. For example, assessing the improvement in water quality after a new wastewater treatment plant is built, or the effect of increased industrial growth and effluent discharges.

A simple but useful modelling approach that may be used in the absence of monitoring data is 'dilution calculations'. In this approach the rate of pollutant loading from point sources in a water body is divided by the streamflow to give a set of estimated pollutant concentrations that may be compared to the standard. Simple dilution calculations assume conservative movement of pollutants. Thus, the use of dilution calculations will tend to be conservative and predict higher than actual concentrations for decaying pollutants. Of course, one could include a best estimate of the effects of decay processes in the dilution model.

Combined runoff and water quality prediction models link stressors (sources of pollutants and pollution) to responses. Stressors include human activities likely to cause impairment, such as the presence of impervious surfaces in a watershed, cultivation of fields close to the stream, over-irrigation of crops with resulting polluted return flows, the discharge of domestic and industrial effluents into water bodies, installing dams and other channelization works, introduction of non-indigenous taxa and over-harvesting of fish. Indirect effects of humans include land cover changes that alter the rates of delivery of water, pollutants and sediment to water bodies.

A review of direct and indirect effects of human activities suggests five major types of environmental stressors:

- alterations in physical habitat
- modifications in the seasonal flow of water
- changes in the food base of the system
- changes in interactions within the stream biota
- release of contaminants (conventional pollutants) (Karr, 1990; NRC, 1992, 2001).

Ideally, models designed to manage water quality should consider all five types of alternative management measures. A broad-based approach that considers these five features provides a more integrative approach to reduce the cause or causes of degradation (NRC, 1992).

Models that relate stressors to responses can be of varying levels of complexity. Sometimes, they are simple qualitative conceptual representations of the relationships among important variables and indicators of those variables, such as the statement 'human activities in a watershed affect water quality, including the condition of the river biota'. More quantitative models can be used to make predictions about the assimilative capacity of a water body, the movement of a pollutant from various point and non-point sources through a watershed, or the effectiveness of certain best management practices.

### 3.1. Model Selection Criteria

Water quality predictive models include both mathematical expressions and expert scientific judgement. They include process-based (mechanistic) models and data-based (statistical) models. The models should link management options to meaningful response variables (such as pollutant sources and water quality standard parameters). They should incorporate the entire 'chain' from stressors to responses. Process-based models should be consistent with scientific theory. Model prediction uncertainty should be reported. This provides decision-makers with estimates of the risks of options. To do this requires prediction error estimates (Chapter 9).

Water quality management models should be appropriate to the complexity of the situation and to the available data. Simple water quality problems can be addressed with simple models, while complex ones may or may not require the use of more complex models. Models requiring large amounts of monitoring data should not be used in situations where such data are unavailable. Models should be flexible enough to allow updates and improvements as appropriate based on new research and monitoring data.

Stakeholders need to accept the models proposed for use in any water quality management study. Given the increasing role of stakeholders in water management decision processes, they need to understand and accept the models being used, at least to the extent they wish to do so. Finally, the cost of maintaining and updating the model over time must be acceptable.

Although predictions are typically made with the aid of mathematical models, there are certainly situations where expert judgement can be just as good. Reliance on professional judgement and simpler models is often acceptable, especially when data are limited.

Highly detailed models require more time and are more expensive to develop and apply. Effective and efficient modelling for water quality management may dictate the use of simpler models. Complex modelling studies should be undertaken only if warranted by the complexity of the management problem. More complex modelling will not necessarily ensure that uncertainty is reduced, and in fact added complexity can compound problems of uncertainty analyses (Chapter 9).

Placing a priority on process description usually leads to the development and use of complex mechanistic models rather than simpler mechanistic or empirical models. In some cases this may result in unnecessarily costly analyses. In addition, physical, chemical and biological processes in terrestrial and aquatic environments are far too complex to be fully represented in even the most complicated models. For water quality management, the primary purpose of modelling should be to support decision-making. The inability to describe all relevant processes completely contributes to the uncertainty in the model predictions.

## 3.2. Model Chains

Many water quality management analyses require the use of a sequence of models, one feeding data into another. For example, consider the sequence or chain of models required for the prediction of fish and shellfish survival as a function of nutrient loadings into an estuary. Of interest to the stakeholders are the conditions of the fish and shellfish. One way to maintain healthy fish and shellfish stocks is to maintain sufficient levels of oxygen in the estuary. The way to do this is to control algae blooms. This in turn requires limits on the nutrient loadings to the estuary that can cause algae blooms, and subsequent dissolved oxygen deficits. The modelling challenge is to link nutrient loading to fish and shellfish survival.

The negative effects of excessive nutrients (e.g. nitrogen) in an estuary are shown in Figure 12.2. Nutrients stimulate the growth of algae. Algae die and accumulate on the bottom, where bacteria consume them. Under calm wind conditions density stratification occurs. Oxygen is depleted in the lower levels of water. Fish and shellfish may die or become weakened and more vulnerable to disease.

A model consisting of a sequence of conditional probabilities can be defined to predict the probability of shellfish and fish abundance based on upstream nutrient loadings into the estuary that might cause problems for fish and shellfish populations. These conditional probabilities can be judgemental, mechanistic and/or statistical. Each conditional probability can be a separate sub-model. Assuming each sub-model can identify a conditional probability distribution, the probability $\Pr\{C \mid N\}$ of a specified amount of carbon, C, given some specified loading of a nutrient, say nitrogen, N, equals the probability $\Pr\{C \mid A\}$ of that given amount of carbon, given a concentration of algae biomass, $A$, times the probability $\Pr\{A \mid N, R\}$ of that concentration of algae biomass given the nitrogen loading, N, and the river flow, $R$, times the probability $\Pr\{R\}$ of the river flow, $R$. In other words:

$$\Pr\{C \mid N\} = \Pr\{C \mid A\}\Pr\{A \mid N, R\}\Pr\{R\} \qquad (12.1)$$

An empirical process-based model of the type to be presented later in this chapter could be used to predict the concentration of algae and the chlorophyll violations on the basis of the river flow and nitrogen loadings. It could similarly predict the production of carbon, on the basis of algae biomass. A seasonal statistical regression model might be used to predict the likelihood of algae blooms based on algal biomass. A cross-system comparison may be made to predict sediment oxygen demand. A relatively simple hydraulic model could be used to

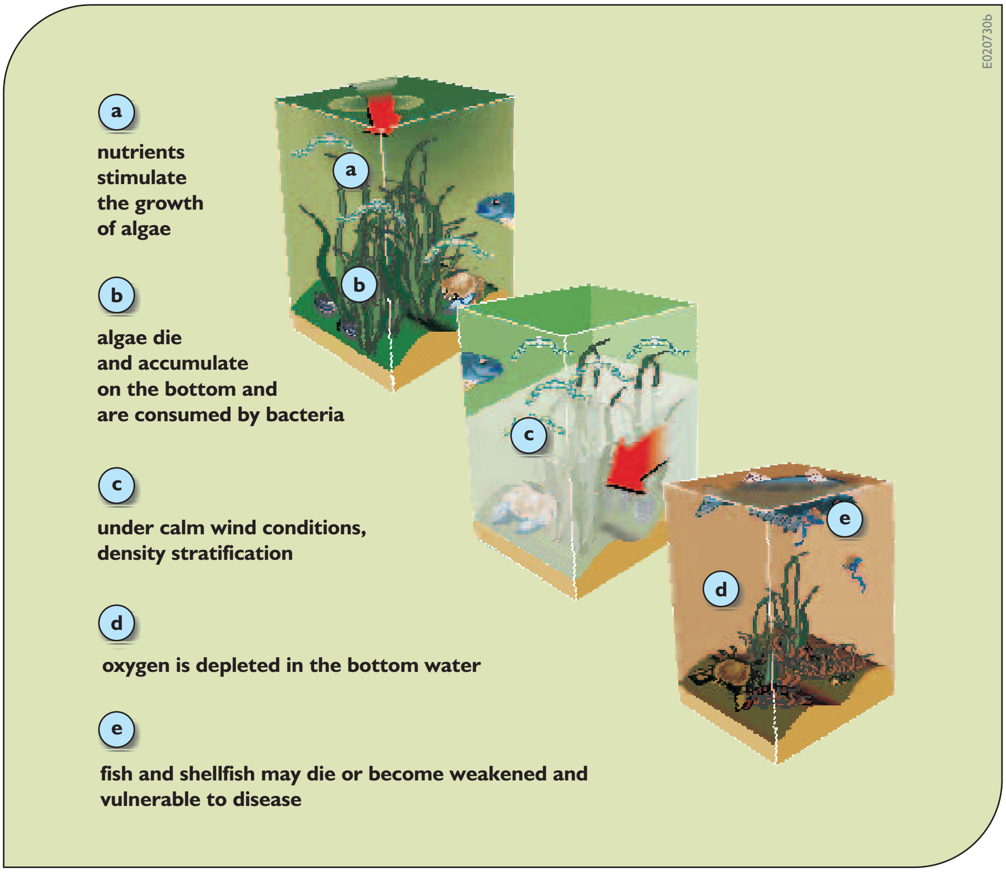



**Figure 12.2.** The negative impacts of excessive nutrients in an estuary (NRC, 2001).

predict the duration of stratification and the frequency of hypoxia, given both the stratification duration and sediment oxygen demand. Expert judgement and fish survival models could be used to predict the shellfish abundance and fishkill and fish health probabilities.

The biological endpoints 'shell-fish survival' and 'number of fishkills', are meaningful indicators to stakeholders and can easily be related to designated water body use. Models and even conditional probabilities assigned to each link of the network in Figure 12.3 can reflect a combination of simple mechanisms, statistical (regression) fitting and expert judgement.

Advances in the mechanistic modelling of aquatic ecosystems have enabled us to include greater process (especially trophic) detail and complexity, as well as to perform dynamic simulations, although mechanistic ecosystem models have not advanced to the point of being able to predict community structure or biotic integrity. In this chapter, only some of the simpler mechanistic models will be introduced. More detail can be found in books solely devoted to water quality modelling (Chapra, 1997; McCutcheon, 1989; Orlob, 1983; Schnoor, 1996; Thomann and Mueller, 1987) as well as the current professional journal literature.

## 3.3. Model Data

Data availability and accuracy are sources of concern in the development and use of models for water quality management. The complexity of models used for water quality management should be compatible with the quantity and quality of available data. The use of complex mechanistic models for water quality prediction in situations with little useful water quality data does not compensate for that lack of data. Model complexity can give the impression of credibility, but this is usually misleading.

It is often preferable to begin with simple models and then, over time, add additional complexity as justified by the collection and analysis of additional data. This strategy makes efficient use of resources. It targets the effort toward information and models that will reduce

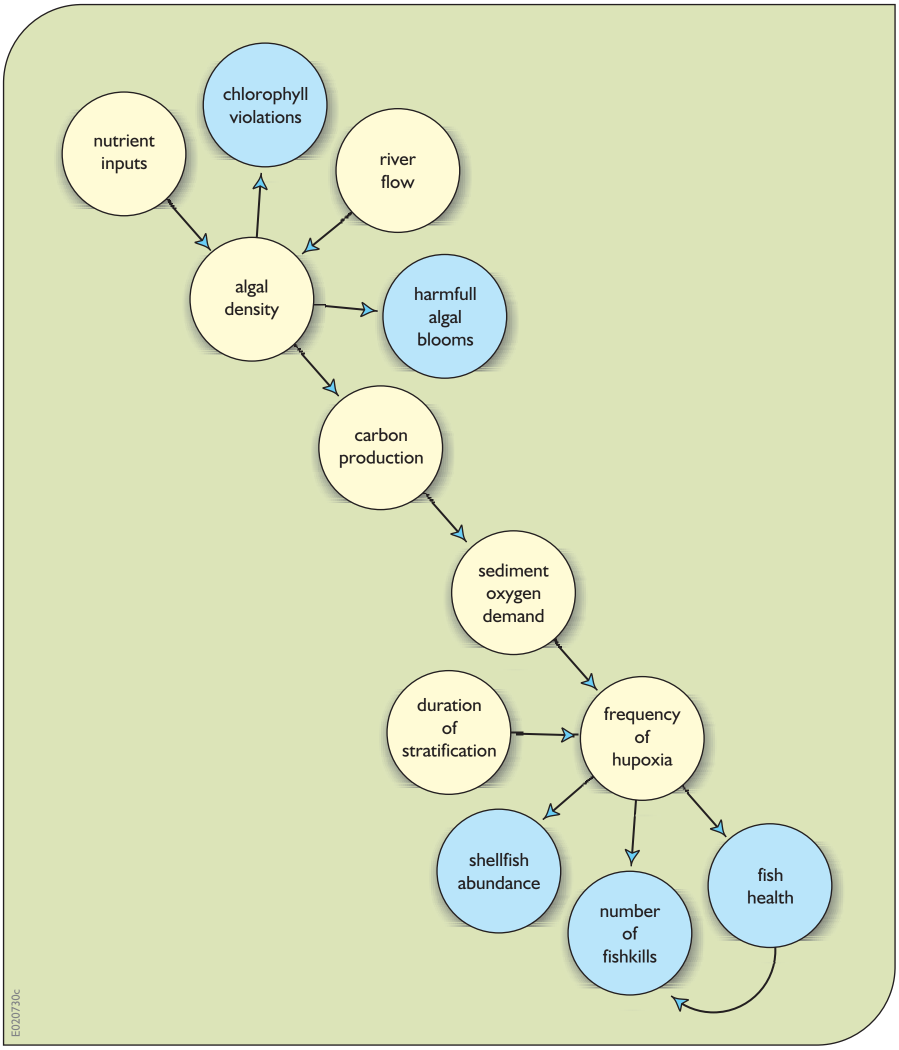



**Figure 12.3.** Cause and effect diagram for estuary eutrophication due to excessive nutrient loadings (Borsuk et al., 2004).

the uncertainty as the analysis proceeds. Models should be selected (simple versus complex) in part on the basis of the data available to support their use.

Water quality models of water bodies receiving pollutant discharges require those pollutant loadings as input data. These pollutant discharges can be from point and non-point sources. Point source discharges are much easier to measure, monitor and estimate than non-point source inputs. Non-point discharge data often come from rainfall–runoff models that attempt to predict the quantity of runoff and its constituent concentrations. The reliability of the predictions from these models is not very good, especially if short time periods (e.g. each day or week) are being simulated. Their average values over longer time periods (e.g. a month or year) tend to be more reliable. This is mainly because the short-term inputs to those models, such as constituent loadings on the land and the rainfall within an area, can vary over space and time within the area and time period being simulated, and are typically not known with any precision. Chapter 13 reviews some of these loading models and their limitations.

# 4. Water Quality Model Processes

Water quality models can be applied to many different types of water system, including streams, rivers, lakes, reservoirs, estuaries, coastal waters and oceans. The models describe the main water quality processes, and typically require the hydrological and constituent inputs (the water flows or volumes and the pollutant loadings). These models include terms for dispersive and/or advective transport depending on the hydrological and hydrodynamic characteristics of the water body, and terms for the biological, chemical and physical reactions among constituents. Advective transport dominates in flowing rivers. Dispersion is the predominant transport phenomenon in estuaries subject to tidal action. Lake-water quality prediction is complicated by the influence of random wind directions and velocities that often affect surface mixing, currents and stratification. For this and other reasons, obtaining reliable quality predictions for lakes is often more difficult than for streams, rivers and estuaries. In coastal waters and oceans, large-scale flow patterns and tide are the most important transport mechanisms.

The development and application of water quality models is both a science and an art. Each model reflects the creativity of its developer, the particular water quality management problems and issues being addressed, the available data for model parameter calibration and verification, the time available for modelling and associated uncertainty, and other considerations. The fact that most, if not all, water quality models cannot accurately predict what actually happens does not detract from their value. Even relatively simple models can help managers understand the real world prototype and estimate at least the relative, if not actual, change in water quality associated with given changes in the inputs resulting from management policies or practices.

### 4.1. Mass-Balance Principles

The basic principle of water quality models is that of mass balance. A water system can be divided into different segments or volume elements, also called 'computational cells'. For each segment or cell, there must be a mass balance for each water quality constituent over time. Most water quality simulation models simulate quality over a consecutive series of discrete time periods, $\Delta t$. Time is divided into discrete intervals $t$ and the flows are assumed constant within each of those time period intervals. For each segment and each time period, the mass balance of a substance in a segment can be defined. Components of the mass balance for a segment include: first, changes by transport ($Tr$) into and out of the segment; second, changes by physical or chemical processes ($P$) occurring within the segment; and third, changes by sources/discharges to or from the segment ($S$).

$$M_i^{t+\Delta t} = M_i^t + \Delta t\left(\frac{\Delta M_i}{\Delta t}\right)_{Tr} + \Delta t\left(\frac{\Delta M_i}{\Delta t}\right)_P + \Delta t\left(\frac{\Delta M_i}{\Delta t}\right)_S \tag{12.2}$$

The mass balance has the following components:

- the mass in computational cell $i$ at the beginning of a time step $t$: $M_i^t$
- the mass in computational cell $i$ at the end of a time step $t$: $M_i^{t+\Delta t}$
- changes in computational cell $i$ by transport: $\left(\frac{\Delta M_i}{\Delta t}\right)_{Tr}$
- changes in computational cell $i$ by physical, (bio)chemical or biological processes: $\left(\frac{\Delta M_i}{\Delta t}\right)_P$
- changes in computational cell $i$ by sources (e.g. wasteloads, river discharges): $\left(\frac{\Delta M_i}{\Delta t}\right)_S$

Changes by transport include both advective and dispersive transport. Advective transport is transport by flowing water. Dispersive transport results from concentration differences. Dispersion in the vertical direction is important if the water column is stratified, and dispersion in the horizontal direction can be in one or two dimensions. Dispersion, as defined here, differs from the physical concept of molecular diffusion as it stands for all transport that is not advective.

Changes by processes include physical processes such as re-aeration and settling, (bio)chemical processes such as adsorption, transformation and denitrification, and biological processes such as primary production and predation on phytoplankton. Water quality processes convert one substance to another.

Changes by sources include the addition of mass by wasteloads and the extraction of mass by intakes. Mass entering over the model boundaries can be considered a source as well. The water flowing into or flowing out of the modelled segment or volume element (the computational cell) is derived from a water quantity (possibly hydrodynamic) model.

To model the transport of substances over space, a water system is divided in small segments or volume elements. The complete ensemble of all the segments or elements is called the *grid* or *schematization*. Each computational cell is defined by its volume and its dimensions in one, two or three directions ($\Delta x$, $\Delta y$, $\Delta z$) depending on the nature of the schematization (1D, 2D or 3D). Note that the cell dimensions $\Delta x$, $\Delta y$ and $\Delta z$ do not have to be equal. The computational cell can have any rectangular shape. A computational cell can share surface areas with other cells, the atmosphere, or the bottom sediment or coast line.

The following sections will look at the transport processes in more detail, defining parameters or variables and their units in terms of mass M, length L and time T.

### 4.1.1. Advective Transport

The advective transport, $T^A_{x_0}$ (M/T), of a constituent at a site $x_0$ is the product of the average water velocity, $v_{x_0}$ (L/T), at that site, the surface or cross-sectional area, $A$ (L$^2$), through which advection takes place at that site, and the average concentration, $C_{x_0}$ (M/L$^3$), of the constituent:

$$T^A_{x_0} = v_{x_0} \times A \times C_{x_0} \tag{12.3}$$

### 4.1.2. Dispersive Transport

The dispersive transport, $T^D_{x_0}$ (M/T), across a surface area is assumed to be proportional to the concentration gradient $\left.\frac{\partial C}{\partial x}\right|_{x=x_0}$ at site $x_0$ times the surface area $A$.

Letting $D_{x_0}$ (L$^2$/T), be the dispersion or diffusion coefficient at site $x_0$:

$$T^D_{x_0} = -D_{x_0} \times A \times \left.\frac{\partial C}{\partial x}\right|_{x=x_0} \tag{12.4}$$

Dispersion is done according to Fick's diffusion law. The minus sign originates from the fact that dispersion causes net transport from higher to lower concentrations, and so in the opposite direction of the concentration gradient. The concentration gradient is the difference of concentrations per unit length, over a very small distance across the cross section:

$$\left.\frac{\partial C}{\partial x}\right|_x = \lim_{\Delta x \to 0} \frac{C_{x+0.5\Delta x} - C_{x-0.5\Delta x}}{\Delta x} \tag{12.5}$$

Dispersion coefficients should be calibrated or be obtained from calculations using turbulence models.

### 4.1.3. Mass Transport by Advection and Dispersion

If the advective and dispersive terms are added and the terms at a second surface at site $x_0 + \Delta x$ are included, a one dimensional equation results:

$$M_i^{t+\Delta t} = M_i^t + \Delta t \times \left( v_{x_0} C_{x_0} - v_{x_0+\Delta x} C_{x_0+\Delta x} - D_{x_0} \left.\frac{\partial C}{\partial x}\right|_{x_0} + D_{x_0+\Delta x} \left.\frac{\partial C}{\partial x}\right|_{x_0+\Delta x} \right) \times A \tag{12.6}$$

or equivalently:

$$M_i^{t+\Delta t} = M_i^t + \Delta t \times \left( Q_{x_0} C_{x_0} - Q_{x_0+\Delta x} C_{x_0+\Delta x} - D_{x_0} A_{x_0} \left.\frac{\partial C}{\partial x}\right|_{x_0} + D_{x_0+\Delta x} A_{x_0+\Delta x} \left.\frac{\partial C}{\partial x}\right|_{x_0+\Delta x} \right) \tag{12.7}$$

where $Q_{x_0}$ (L$^3$/T) is the flow at site $x_0$.

If the previous equation is divided by the volume and the time interval $\Delta t$, then the following equation results in one dimension:

$$\frac{C_i^{t+\Delta t} - C_i^t}{\Delta t} = \frac{D_{x_0+\Delta x} \left.\frac{\partial C}{\partial x}\right|_{x_0+\Delta x} - D_{x_0} \left.\frac{\partial C}{\partial x}\right|_{x_0}}{\Delta x} + \frac{v_{x_0} C_{x_0} - v_{x_0+\Delta x} C_{x_0+\Delta x}}{\Delta x} \tag{12.8}$$

Taking the asymptotic limit $\Delta t \to 0$ and $\Delta x \to 0$, the advection–diffusion equation for one dimension results:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left( D \frac{\partial C}{\partial x} \right) - \frac{\partial}{\partial x}(vC) \tag{12.9}$$

The finite volume method for transport is a computational method of solving the advection–diffusion equation. The accuracy of the method will be related to the size of $\Delta x$, $A$ ($A = \Delta y \Delta z$) and $\Delta t$.

By adding terms for transport in the y and z-direction, a three-dimensional model is obtained. Taking the asymptotic limit again will lead to a three-dimensional advection–diffusion equation

$$\frac{\partial C}{\partial t} = D_x \frac{\partial^2 C}{\partial x^2} - v_x \frac{\partial C}{\partial x} + D_y \frac{\partial^2 C}{\partial y^2} - v_y \frac{\partial C}{\partial y} + D_z \frac{\partial^2 C}{\partial z^2} + v_z \frac{\partial C}{\partial z} + S + f_R(C, t) \quad (12.10)$$

with dispersion coefficients $D_j$ defined for each direction. If source terms '$S$' and '$f_R$' are added as shown in the equation above, the so-called advection–diffusion reaction equation emerges. The additional terms represent:

- Discharges or 'wasteloads' ($S$): these source terms are additional inflows of water or mass. As many source terms as required may be added to Equation 12.10. These could include small rivers, discharges of industries, sewage treatment plants, small wasteload outfalls and so on.
- Reaction terms or 'processes' ($f_R$).

Processes can be split into physical and other processes. Examples of physical processes are:

- settling of suspended particulate matter
- water movement not affecting substances, like evaporation
- volatilization of the substance itself at the water surface.

Examples of other processes are:

- biochemical conversions like ammonia and oxygen forming nitrite
- growth of algae (primary production)
- predation by other animals
- chemical reactions.

These processes are described in more detail in the remaining parts of this section.

The expression $D(\partial C/\partial X) - vC$ in Equation 12.9, multiplied by the area $A$, is termed the total flux (M/T). Flux due to dispersion, $DA(\partial C/\partial X)$, is assumed to be proportional to the concentration gradient over distance. Constituents are transferred by dispersion from higher concentration zones to lower ones. The coefficient of dispersion $D$ ($L^2$/T) depends on the amplitude and frequency of the tide, if applicable, as well as upon the turbulence of the water body. It is common practice to include in this dispersion parameter everything affecting the distribution of $C$ other than advection. The term $vAC$ is the advective flux caused by the movement of water containing the constituent concentration $C$ (M/$L^3$) at a velocity rate $v$ (L/T) across a cross-sectional area $A$ ($L^2$).

The relative importance of dispersion and advection depends on the degree of detail with which the velocity field is defined. A good spatial and temporal description of the velocity field within which the constituent is being distributed will reduce the importance of the dispersion term. Less precise descriptions of the velocity field, such as averaging across irregular cross sections or approximating transients by steady flows, may lead to a dominance of the dispersion term.

Many of the reactions affecting the decrease or increase of constituent concentrations are often represented by first-order kinetics that assume the reaction rates are proportional to the constituent concentration. While higher-order kinetics may be more correct in certain situations, predictions of constituent concentrations based on first-order kinetics have often been found to be acceptable for natural aquatic systems.

## 4.2. Steady-State Models

A steady state means no change over time. If we consider a water body, for example a river, this means there are no changes in the concentrations with time. In this case the left hand side of Equation 12.9, $\partial C/\partial t$, equals 0. Assume the only sink is the natural decay of the constituent defined as $kC$ where $k$, ($T^{-1}$), is the decay rate coefficient or constant. Now Equation 12.9 becomes

$$0 = D\,\partial^2 C/\partial X^2 - v\partial C/\partial X - kC \quad (12.11)$$

Equation 12.11 can be integrated, since river reach parameters $A$, $D$, $k$, $v$, and $Q$ are assumed constant. For a constant loading, $W_C$ (M/T) at site $X = 0$, the concentration $C$ at any distance $X$ will equal

$$C(X) = \begin{cases} (W_C/Qm)\exp[(v/2D)(1+m)X] & X \le 0 \\ (W_C/Qm)\exp[(v/2D)(1-m)X] & X \ge 0 \end{cases} \quad (12.12)$$

where

$$m = (1 + (4kD/v^2))^{1/2} \quad (12.13)$$

Note from Equation 12.12 that the parameter $m$ is always equal to or greater than 1 and that the exponent of $e$ is always negative. Hence, as the distance $X$ increases in magnitude, either in the positive or negative direction, the concentration $C(X)$ will decrease if $k > 0$. The maximum concentration $C$ occurs at $X = 0$ and is $W_C/Qm$.

$$C(0) = W_C/Qm \tag{12.14}$$

These equations are plotted in Figure 12.4.

In flowing rivers not under the influence of tidal actions the dispersion is usually small. Assuming the dispersion coefficient $D$ is 0, the parameter $m$ defined by Equation 12.13 is 1. Hence, when $D = 0$, the maximum concentration at $X = 0$ is $W_C/Q$.

$$C(0) = W_C/Q \quad \text{if } D = 0. \tag{12.15}$$

Assuming $D = 0$ and $v$, $Q$ and $k > 0$, Equation 12.12 becomes

$$C(X) = \begin{cases} 0 & X \leq 0 \\ (W_C/Q)\exp[-kX/v] & X \geq 0 \end{cases} \tag{12.16}$$

The above equation for $X > 0$ can be derived from Equations 12.12 and 12.13 by noting that the term $(1 - m)$ equals $(1 - m)(1 + m)/(1 + m) = (1 - m^2)/2$ when $D$ is 0. Thus when $D$ is 0 the expression $v/2D)(1 - m)X$ in Equation 12.12 becomes $-kX/v$. The term $X/v$ is sometimes denoted as a single variable representing the time of flow: the time flow $Q$ takes to travel from site $X = 0$ to some other downstream site for a distance of $X$.

As rivers approach the sea, the dispersion coefficient $D$ increases and the net downstream velocity $v$ decreases. Because the flow $Q$ equals the cross-sectional area $A$ times the velocity $v$, $Q = Av$, and since the parameter $m$ can be defined as $(v^2 + 4kD)^{1/2}/v$, then as the velocity $v$ approaches 0, the term $Qm = Av(v^2 + 4kD)^{1/2}/v$ approaches $2A(kD)^{1/2}$. The exponent $vX(1 \pm m)/2D$ in Equation 12.12 approaches $\pm X(k/D)^{1/2}$.

Hence for small velocities, Equation 12.12 becomes

$$C(X) = \begin{cases} (W_C/2A(kD)^{1/2})\exp[+X(k/D)^{1/2}] & X \leq 0 \\ (W_C/2A(kD)^{1/2})\exp[-X(k/D)^{1/2}] & X \geq 0 \end{cases} \tag{12.17}$$

Here, dispersion is much more important than advective transport and the concentration profile approaches a symmetric distribution, as shown in Figure 12.4, about the point of discharge at $X = 0$.

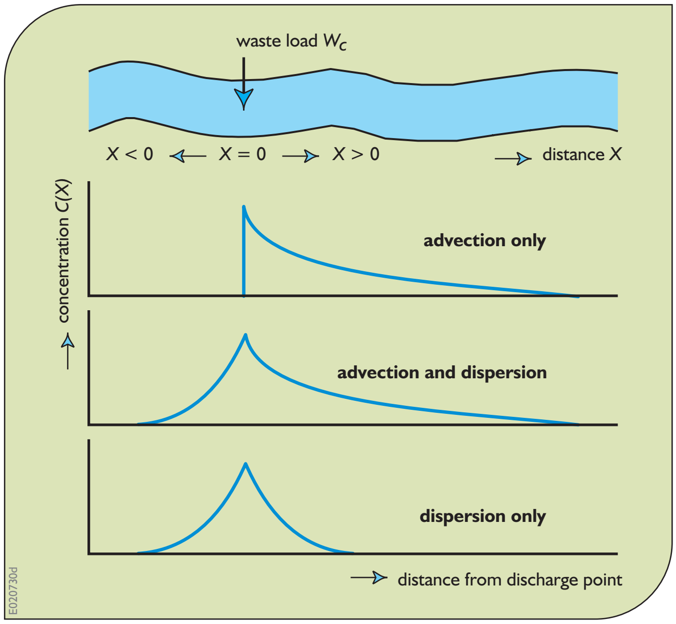
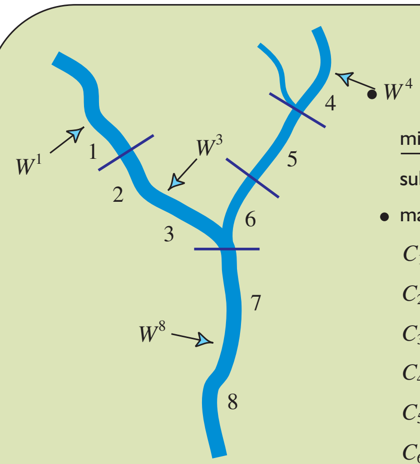



**Figure 12.4. Constituent concentration distribution along a river or estuary resulting from a constant discharge of that constituent at a single point source in that river or estuary.**

Water quality management models are often used to assess the effect of pollutant loadings on ambient waters and to compare the results with specific water quality standards. The above steady-state equations can be used to construct such a model for estimating the wastewater removal efficiencies required at each wastewater discharge site that will result in an ambient stream quality that meets the standards along a stream or river.

Figure 12.5 shows a schematic of a river into which wastewater containing constituent $C$ is being discharged at four sites. Assume that maximum allowable concentrations of the constituent $C$ are specified at each of those discharge sites. To estimate the necessary reduction in these discharges, the river must be divided into approximately homogenous reaches. Each reach can be characterized by constant values of the cross-sectional area, $A$, dispersion coefficient, $D$, constituent decay rate constant, $k$, and velocity, $v$, associated with some 'design' flow and temperature conditions. These parameter values and the length, $X$, of each reach can differ; hence, the subscript index $i$ will be used to denote the particular parameter values for the particular reach. These reaches are shown in Figure 12.5.




minimize $\Sigma_i \text{Cost}_i(R_i)$

subject to:

- mass balances:

$$C_1 = W^1(1\text{-}R_1)/(Qm)_1 + C_2\, TB_1$$
$$C_2 = [C_1 TF_1 (Qm)_1 + C_3 TB_2 (Qm)_2] \,/\, (Qm)_2$$
$$C_3 = [W^3(1\text{-}R_3) + C_2 TF_2 (Qm)_2 + C_7 TB_3 (Qm)_3] \,/\, (Qm)_3$$
$$C_4 = [W^4(1\text{-}R_4)/(Qm)_4 + C_5 TB_4$$
$$C_5 = [C_4 TF_4 (Qm)_4 + C_6 TB_5 (Qm)_5] \,/\, (Qm)_5$$
$$C_6 = [C_5 TF_5 (Qm)_5 + C_7 TB_6 (Qm)_6] \,/\, (Qm)_6$$
$$C_7 = [C_3 TF_3 (Qm)_3 + C_6 TF_6 (Qm)_6 + CB_8 TB_7 (Qm)_7] \,/\, (Qm)_7$$
$$C_8 = [W^8(1\text{-}R_8) + C_7 TF_7 (Qm)_7 ] \,/\, (Qm)_8$$

- quality standards and maximum removal efficiencies:

$$C_i \le C_i^{\max} \;\; \forall i \,;\;\; R_i \le R_i^{\max} \;\; \forall i$$

**Figure 12.5.** Optimization model for finding constituent removal efficiencies, *Ri*, at each discharge site *i* that result in meeting stream quality standards, $Ci^{max}$, at least total cost.

In Figure 12.5 each variable $C_i$ represents the constituent concentration at the beginning of reach $i$. The flows $Q$ represent the design flow conditions. For each reach $i$ the product $(Q_i m_i)$ is represented by $(Qm)_i$. The downstream (forward) transfer coefficient, $TF_i$, equals the applicable part of Equation 12.12,

$$TF_i = \exp[(v/2D)(1 - m)X] \tag{12.18}$$

as does the upstream (backward) transfer coefficient, $TB_i$.

$$TB_i = \exp[(v/2D)(1 + m)X] \tag{12.19}$$

The parameter $m$ is defined by Equation 12.13.

Solving a model such as the one shown in Figure 12.5 does not mean that the least-cost wasteload allocation plan will be implemented, but least cost solutions can identify the additional costs of other imposed constraints, for example, to ensure equity or extra safety. Models like this can be used to identify the cost-quality tradeoffs inherent in any water quality management programme. Other than economic objectives can also be used to obtain other tradeoffs.

The model in Figure 12.5 incorporates both advection and dispersion. If upstream dispersion under design streamflow conditions is not significant in some reaches, then the upstream (backward) transfer coefficients, $TB_i$, for those reaches $i$ will equal 0.

### 4.3. Design Streamflows for Water Quality

In streams and rivers, the water quality may vary significantly, depending on the water flow. If wasteload discharges are fairly constant, a high flow serves to dilute the waste concentration, while where there is a low flow concentrations may become undesirably high. It is therefore common practice to pick a low-flow condition for judging whether or not ambient water quality standards are being met. This can also be seen from Equations 12.12, 12.14, 12.15, and 12.16. This often is the basis for the assumption that the smaller (or more critical) the design flow, the more likely it is that the stream quality standards will be met. This is not always the case, however.

Different regions of the world use different design low-flow conditions. One example of such a design flow, which is used in parts of North America, is the minimum seven-day average flow expected once in ten years on average. Each year the lowest seven-day average flow is determined, as shown in Figure 12.6. The sum of each of the 365 sequences of seven average daily flows is divided by seven, and the minimum value is selected. This is the minimum annual average seven-day flow.

These minimum seven-day average flows for each year of record define a probability distribution whose

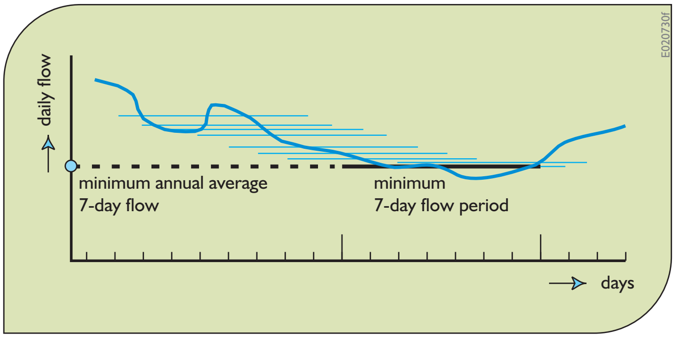



**Figure 12.6. Portion of annual flow time series showing low flows and the calculation of average seven and fourteen-day flows.**

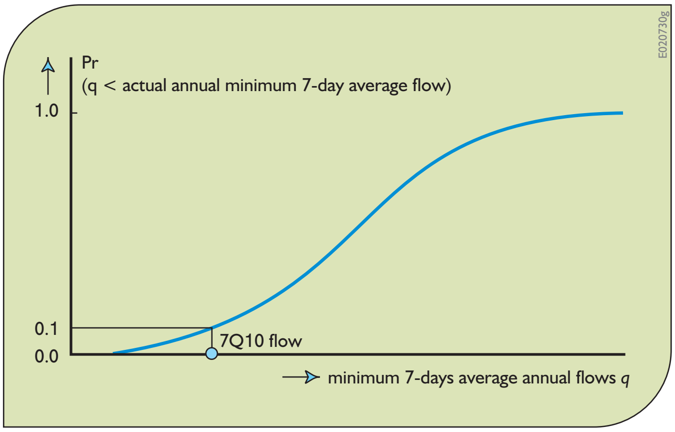



**Figure 12.7. Determining the minimum seven-day annual average flow expected once in ten years, designated 7Q10, from the cumulative probability distribution of annual minimum seven-day average flows.**

cumulative probabilities can be plotted. As illustrated in Figure 12.7, the particular flow on the cumulative distribution that has a 90% chance of being exceeded is the design flow. It is the minimum annual average seven-day flow expected once in ten years. This flow is commonly called the 7Q10 flow. Analyses have shown that this daily design flow is exceeded about 99% of the time in regions where it is used (NRC, 2001). This means that there is on average only a 1% chance that any daily flow will be less than this 7Q10 flow.

Consider now any one of the river reaches shown in Figure 12.5. Assume an initial amount of constituent mass, $M$, exists at the beginning of the reach. As the reach volume, $Q\Delta t$, increases due to the inflow of less polluted water, the initial concentration, $M/Q\Delta t$, will decrease. However, the flow velocity will increase, and thus the time, $\Delta t$, it takes to transport the constituent mass to the end of that reach will decrease. This means less time for the decay of the constituent. Thus, wasteload allocations that meet ambient water quality standards during low-flow conditions may not meet them under higher-flow conditions, which are observed much more frequently. Figure 12.8 illustrates how this might happen. This does not suggest low flows should not be considered when allocating wasteloads, but rather that a simulation of water quality concentrations over varying flow conditions may show that higher-flow conditions at some sites are even more critical and more frequent than are the low-flow conditions.

Figure 12.8 shows that for a fixed mass of pollutant at $X = 0$, under low-flow conditions the more restrictive (lower) maximum pollutant concentration standard in the downstream portion of the river is met, but that same standard is violated under more frequent higher-flow conditions.

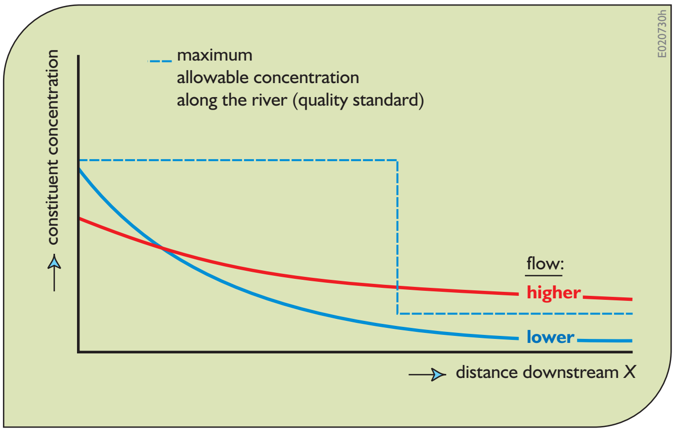



**Figure 12.8. Increasing the streamflows decreases initial concentrations but may increase downstream concentrations.**

## 4.4. Temperature

Temperature affects almost all water quality processes taking place in water bodies. For this reason, it may be important to model temperature when it may vary substantially over the period of interest, or when the discharge of heat into water bodies is to be managed.

Temperature models are based on a heat balance in the water body. A heat balance takes into account the sources and sinks of heat. The main sources of heat in a water body are short-wave solar radiation, long-wave atmospheric radiation, conduction of heat from the atmosphere to the water and direct heat inputs. The main sinks of heat are long-wave radiation emitted by the water, evaporation and conduction from the water to the atmosphere. Unfortunately, a model with all the sources and sinks of heat requires measurements of a number of variables and coefficients that are not always readily available.

One temperature predictor is the simplified model that assumes an equilibrium temperature $T_e$ (°C) will be reached under steady-state meteorological conditions. The temperature mass balance in a volume segment depends on the water density $\rho$ (g/cm$^3$), the heat capacity of water, $c_p$ (cal/g/°C), and the water depth $h$ (cm). The net heat input, $K_H(T_e - T)$ (cal/cm$^2$/day), is assumed to be proportional to the difference of the actual temperature, $T$, and the equilibrium temperature, $T_e$ (°C).

$$dT/dt = K_H(T_e - T)/\rho c_p h \tag{12.20}$$

The overall heat exchange coefficient, $K_H$ (cal/cm$^2$/day/°C), is determined in units of Watts/m$^2$/°C (1 cal/cm$^2$/day/°C = 0.4840 Watts/m$^2$/°C) from empirical relationships that include wind velocity, dew point temperature and actual temperature $T$ (°C) (Thomann and Mueller, 1987).

The equilibrium temperature, $T_e$, is obtained from another empirical relationship involving the overall heat exchange coefficient, $K_H$, the dew point temperature, $T_d$, and the short-wave solar radiation, $H_s$ (cal/cm$^2$/day),

$$T_e = T_d + (H_s/K_H) \tag{12.21}$$

This model simplifies the mathematical relationships of a complete heat balance and requires less data.

## 4.5. Sources and Sinks

Sources and sinks of pollutants include wasteloads, and the physical and biochemical processes that alter those wasteloads. External inputs of each constituent would have the form $W/Q\Delta t$ or $W/(A_X\Delta X)$ where $W$ (M/T) is the loading rate of the constituent and $Q\Delta t$ or $A_X\Delta X$ (L$^3$) represents the volume of water into which the mass of waste $W$ is discharged. Constituent growth and decay processes are discussed in the remaining parts of this Section 4.

## 4.6. First-Order Constituents

The first-order models are commonly used to predict water quality constituent decay or growth. They can represent constituent reactions such as decay or growth in situations where the time rate of change ($dC/dt$) in the concentration $C$ of the constituent, say organic matter that creates a biochemical oxygen demand ($BOD$), is proportional to the concentration of either the same or another constituent concentration. The temperature-dependent proportionality constant $k_c$ (1/day) is called a rate coefficient or constant. In general, if the rate of change in some constituent concentration $C_j$ is proportional to the concentration $C_i$, of constituent $i$, then

$$dC_j/dt = a_{ij} k_i \theta_i^{(T-20)} C_i \tag{12.22}$$

where $\theta_i$ is the temperature correction coefficient for $k_i$ at 20 °C and $T$ is the temperature in °C. The parameter $a_{ij}$ is the grams of $C_j$ produced ($a_{ij} > 0$) or consumed ($a_{ij} < 0$) per gram $C_i$. For the prediction of $BOD$ concentration over time, $C_i = C_j = BOD$ and $a_{ij} = a_{BOD} = -1$ in Equation 12.22. Conservative substances, such as salt, will have a decay rate constant $k$ of 0. The concentration of conservative substances depends only on the amount of water, that is, dilution.

The typical values for the rate coefficients $k_c$ and temperature coefficients $\theta_i$ of some constituents $C$ are in Table 12.1. For bacteria, the first-order decay rate ($k_B$) can also be expressed in terms of the time to reach 90% mortality ($t_{90}$, days). The relationship between these coefficients is given by $k_B = 2.3/t_{90}$.

## 4.7. Dissolved Oxygen

Dissolved oxygen ($DO$) concentration is a common indicator of the health of the aquatic ecosystem. $DO$ was originally modelled in the Ohio River (US) by Streeter and Phelps (1925). Since then a number of modifications and extensions of the model have been made relating to the number of sinks and sources of $DO$ being considered, and how processes involving the nitrogen cycle and phytoplankton are being modelled, as illustrated in Figure 12.9.

The sources of $DO$ in a water body include re-aeration from the atmosphere, photosynthetic oxygen production from aquatic plants, denitrification and $DO$ inputs. The

| constituent | rate constant *k* | | units |
|---|---|---|---|
| total coliform bacteria (freshwater) | 1.0-5.5 | - a | 1/day |
| total coliform bacteria (sediments) | 0.14-0.21 | - a | 1/day |
| total coliform bacteria (seawater) | 0.7-3.0 | - a | 1/day |
| fecal coliform bacteria (seawater) | 37-110 | - a | 1/day |
| *BOD* (no treatment) | 0.3-0.4 | - a | 1/day |
| *BOD* (activated sludge treatment) | 0.05-0.1 | - a | 1/day |
| carbofuran | 0.03 | - b | 1/day |
| *DDT* | 0.0-0.10 | - b | 1/day |
| *PCB* | 0.0-0.007 | - b | 1/day |
| pentachlorophenol | 0.0-33.6 | - b | 1/day |

| constituent | θ | | units |
|---|---|---|---|
| coliform bacteria (freshwater) | 1.07 | - b | — |
| coliform bacteria (saltwater) | 1.10 | - b | — |
| *BOD* | 1.04 | - a | --- |

a - Thomann and Mueller (1987)   b - Schnoor (1996)

E020913a

**Table 12.1.** Typical values of the first-order decay rate, *k*, and the temperature correction factor, $\theta$, for some constituents.

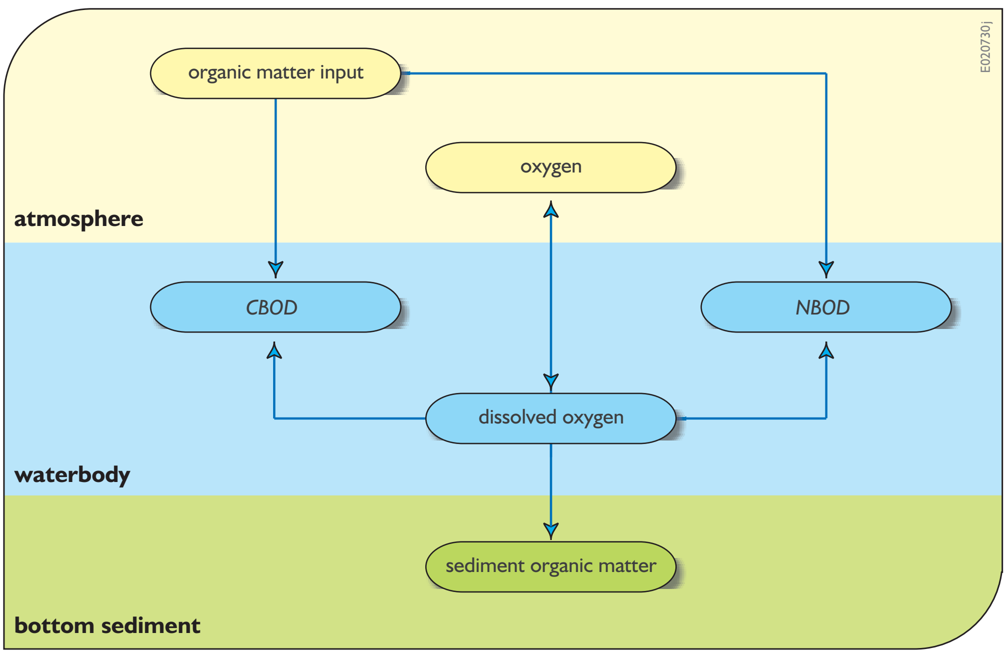



**Figure 12.9.** The dissolved oxygen interactions in a water body, showing the decay (satisfaction) of carbonaceous, nitrogenous and sediment oxygen demands and water body re-aeration or deaeration (if supersaturation occurs at the air–water interface).

sinks include oxidation of carbonaceous and nitrogenous material, sediment oxygen demand and respiration by aquatic plants.

$$\Delta O_2/\Delta t = \text{loads} + \text{transport} + \text{re-aeration} + \text{net primary production} + \text{denitrification} - \text{mineralization} - \text{nitrification} - SOD$$

The rate of re-aeration is assumed to be proportional to the difference between the saturation concentration, $DO_{sat}$ (mg/l), and the concentration of dissolved oxygen, $DO$ (mg/l). The proportionality coefficient is the re-aeration rate $k_r$ (1/day), defined at temperature $T = 20\,°C$ , which can be corrected for any temperature $T$ with the coefficient $\theta_r^{(T-20)}$. The value of this temperature correction coefficient, $\theta$, depends on the mixing condition of the water body. Values generally range from 1.005 to 1.030. In practice, a value of 1.024 is often used (Thomann and Mueller, 1987). The re-aeration rate constant is a sensitive parameter. There have been numerous equations developed to define this rate constant. Table 12.2 lists some of them.

The saturation concentration, $DO_{sat}$, of oxygen in water is a function of the water temperature and salinity (chloride concentration, Cl (g/m$^3$)), and can be approximated by

$$DO_{sat} = \{14.652 - 0.41022T + (0.089392T)^2 - (0.042685T)^3\}\{1 - (Cl/100{,}000)\} \qquad (12.23a)$$

Elmore and Hayes (1960) derived an analytical expression for the $DO$ saturation concentration, $DO_{sat}$ (mg/l), as a function of temperature ($T$, °C):

$$DO_{sat} = 14.652 - 0.41022T + 0.007991T^2 - 0.000077774T^3 \qquad (12.23b)$$

Fitting a second-order polynomial curve to the data presented in Chapra (1997) results in:

$$DO_{sat} = 14.407 - 0.3369T + 0.0035T^2 \qquad (12.23c)$$

as is shown in Figure 12.10.

Because photosynthesis occurs during daylight hours, photosynthetic oxygen production follows a cyclic, diurnal, pattern in water. During the day, oxygen concentrations in water are high and can even become supersaturated, i.e. concentrations exceeding the saturation concentration. At night, the concentrations drop due to respiration and other oxygen-consuming processes.

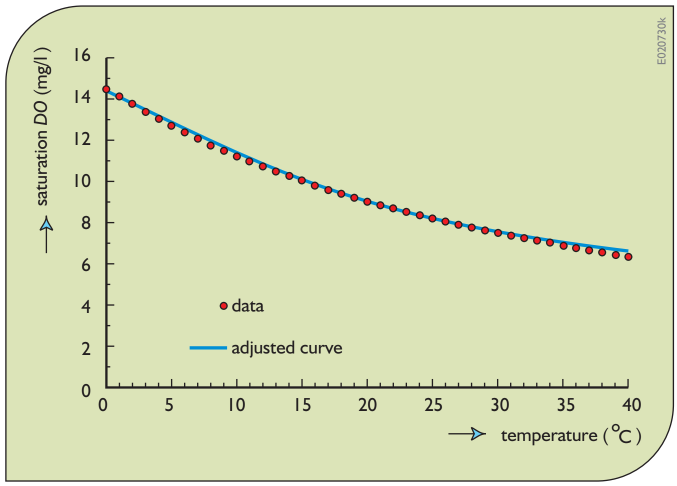



**Figure 12.10.** Fitted curve to the saturation dissolved oxygen concentration (mg/l) as a function of temperature (°C).

**Table 12.2.** Some equations for defining the re-aeration rate constant, $k_r$ (day$^{-1}$).

E020913b

**units** **water and wind velocity (m/s)** **water depth (m)**

$k_r$ = mass transport coefficient for reaeration (m/day) / (water depth)

= 5.026 (water velocity)$^{0.969}$ / (water depth)$^{1.673}$ (Churchill, 1962)

= 3.95 (water velocity)$^{0.5}$ / (water depth)$^{1.5}$ (O'Connor and Dobbiens, 1958)

= (scale factor) 3.95 (water velocity)$^{0.5}$ / (water depth)$^{1.5}$

= 5.344 (water velocity)$^{0.670}$ / (water depth)$^{1.85}$ (Owens, Edwards, Gibb, 1964)

= 5.13 (water velocity) / (water depth)$^{1.333}$ (Langbien, Durum, 1967)

= { 0.065 (wind velocity)$^2$ + 3.86 [(water velocity) / (water depth)]$^{0.5}$ } / (water depth) (van Pagee 1978, Delvigne 1980)

One can distinguish between the biochemical oxygen demand from carbonaceous organic matter (*CBOD*, mg/l) in the water, and that from nitrogenous organic matter (*NBOD*, mg/l) in the water. There is also the oxygen demand from carbonaceous and nitrogenous organic matter in the sediments (*SOD*, mg/l/day). These oxygen demands are typically modelled as first-order decay reactions with decay rate constants $k_{CBOD}$ (1/day) for *CBOD* and $k_{NBOD}$ (1/day) for *NBOD*. These rate constants vary with temperature, hence they are typically defined for 20 °C. The decay rates are corrected for temperatures other than 20 °C using temperature coefficients $\theta_{CBOD}$ and $\theta_{NBOD}$, respectively.

The sediment oxygen demand *SOD* (mg/l/day) is usually expressed as a zero-order reaction, that is, a constant demand. One important feature in modelling *NBOD* is ensuring the appropriate time lag between when it is discharged into a water body and when the oxygen demand is observed. This lag is in part a function of the level of treatment in the wastewater treatment plant.

The dissolved oxygen (*DO*) model with *CBOD*, *NBOD* and *SOD* is

$$\begin{aligned} \mathrm{d}DO/\mathrm{d}t = & -k_{CBOD}\theta_{CBOD}^{(T-20)}CBOD \\ & -k_{NBOD}\theta_{NBOD}^{(T-20)}NBOD \\ & +k_r\theta_r^{(T-20)}(DO_{\mathrm{sat}}-DO)-SOD \end{aligned} \qquad (12.24)$$

$$\mathrm{d}CBOD/\mathrm{d}t = -k_{CBOD}\theta_{CBOD}^{(T-20)}CBOD \qquad (12.25)$$

$$\mathrm{d}NBOD/\mathrm{d}t = -k_{NBOD}\theta_{NBOD}^{(T-20)}NBOD \qquad (12.26)$$

The mean and range values for coefficients included in these dissolved oxygen models are shown in Table 12.3.

## 4.8. Nutrients and Eutrophication

Eutrophication is the progressive process of nutrient enrichment of water systems. The increase in nutrients leads to an increase in the productivity of the water system, which may result in an excessive increase in the biomass of algae or other primary producers, such as macrophytes or duck weed. When it is visible on the surface of the water it is called an algae bloom. Excessive algal biomass could affect the water quality, especially if it causes anaerobic conditions and thus impairs the drinking, recreational and ecological uses.

The eutrophication component of the model relates the concentration of nutrients and the algal biomass. Taking the example shown in Figure 12.11, consider the growth of algae *A* (mg/l), depending on phosphate phosphorus, *P* (mg/l), and nitrite/nitrate nitrogen, $N_n$ mg/l), as the limiting nutrients. There could be other limiting nutrients or other conditions as well, but here consider only these two. If either of these two nutrients is absent, the algae cannot grow, regardless of the abundance of the other nutrient. The uptake of the more abundant nutrient will not occur.

To account for this, algal growth is commonly modelled as a Michaelis–Menten multiplicative effect; in other words, the nutrients have a synergistic effect. Model parameters include a maximum algal growth rate $\mu$ (1/day) times the fraction of a day, $f_d$, that rate applies (Figure 12.12), the half saturation constants $K_P$ and $K_N$ (mg/l) (shown as $K_C$ in Figure 12.13) for phosphate and nitrate, respectively, and a combined algal respiration and specific death rate constant *e* (1/day) that creates an oxygen demand. The uptake of phosphate, ammonia and nitrite/nitrate by algae is assumed to occur in proportion to their contents in the algae biomass. Define these proportions as $a_P$, $a_A$ and $a_N$, respectively.

In addition to the above parameters, one needs to know the amounts of oxygen consumed in the oxidation of organic phosphorus, $P_o$, and the amounts of oxygen produced by photosynthesis and consumed by respiration. In the model below, some average values have been assumed. Also assumed are constant temperature correction factors for all processes pertaining to any individual constituent. This reduces the number of parameters needed, but is not necessarily realistic. Clearly other processes as well as other parameters could be added, but the purpose here is to illustrate how these models are developed. Users of water quality simulation programs will appreciate the many different assumptions that can be made and the large amount of parameters associated with most of them.

The source and sink terms of the relatively simple eutrophication model shown in Figure 12.11 can be written as follows:

For algae biomass:

$$\begin{aligned} \mathrm{d}A/\mathrm{d}t = & \mu f_d\theta_A^{(T-20)}[P/(P+K_P)][N_n/(N_n+K_N)] \\ & \times A - e\theta_A^{(T-20)}A \end{aligned} \qquad (12.27)$$

For organic phosphorus:

$$\mathrm{d}P_o/\mathrm{d}t = -k_{op}\theta_{op}^{(T-20)}P_o \qquad (12.28)$$

**Table 12.3.** Typical values of parameters used in the dissolved oxygen models.

E020913c

| parameter | value | | units |
|---|---|---|---|
| $k_r$, slow, deep rivers | 0.1-0.4 | - a | 1/day |
| $k_r$, typical conditions | 0.4-1.5 | - a | 1/day |
| $k_r$, swift, deep rivers | 1.5-4.0 | - a | 1/day |
| $k_r$, swift, shallow rivers | 4.0-10.0 | - a | 1/day |
| $k_{CBOD}$, untreated discharges | 0.35 (0.20-0.50) | - b | 1/day |
| $k_{CBOD}$, primary treatment | 0.20 (0.10-0.30) | - b | 1/day |
| $k_{CBOD}$, activated sludge | 0.075 (0.05-0.10) | - b | 1/day |
| $\theta_{CBOD}$ | 1.04 | - a | — |
| | 1.047 | - a | — |
| | 1.04 (1.02-1.09) | - c | — |
| $\theta_r$ | 1.024 (1.005-1.030) | - c | --- |

| sediment oxygen demand * | value | | units |
|---|---|---|---|
| municipal sludge (outfall vicinity) | 4 (2-10) | - c d | g $O_2$ / $m^2$ / day |
| municipal sewage sludge | 1.5 (1-2) | - c d | g $O_2$ / $m^2$ / day |
| sandy bottom | 0.5 (0.2-1.0) | - c d | g $O_2$ / $m^2$ / day |
| mineral soils | 0.07 (0.05-0.1) | - c d | g $O_2$ / $m^2$ / day |
| natural to low pollution | 0.1-10.0 | - a | g $O_2$ / $m^2$ / day |
| moderate to heavy pollution | 5-10 | - a | g $O_2$ / $m^2$ / day |

a - Schnoor (1996) c - Thomann and Mueller (1987)
b - Chapra (1997) d - Bowie et al. (1985)

* value has to be divided by the water height (m)

For phosphate phosphorus:

$$dP/dt = -\mu f_d \theta_A^{(T-20)}[P/(P+K_P)] \times [N_n/(N_n+K_N)]a_P A \quad (12.29)$$

For organic nitrogen:

$$dN_o/dt = -k_{on}\theta_{on}^{(T-20)}N_o \quad (12.30)$$

For ammonia-nitrogen:

$$dN_a/dt = -\mu f_d \theta_A^{(T-20)}[P/(P+K_P)] \times [N_n/(N_n+K_N)]a_A A + k_{on}\theta_{on}^{(T-20)}N_o - k_a\theta_a^{(T-20)}N_a \quad (12.31)$$

For nitrate-nitrogen:

$$dN_n/dt = -\mu f_d \theta_A^{(T-20)}[P/(P+K_p)] \times [N_n/(N_n+K_N)]a_N A + k_a\theta_a^{(T-20)}N_a - k_n\theta_n^{(T-20)}N_n \quad (12.32)$$

For dissolved oxygen:

$$dDO/dt = -k_{CBOD}\theta_{CBOD}^{(T-20)}CBOD - 4.57k_a\theta_a^{(T-20)}N_a - 2k_{op}\theta_{op}^{(T-20)}P_o + (1.5\mu f_d - 2e)\theta_A^{(T-20)}A + k_r\theta_r^{(T-20)}(DO_{sat} - DO) - SOD \quad (12.33)$$

WATER RESOURCES SYSTEMS PLANNING AND MANAGEMENT – ISBN 92-3-103998-9 – © UNESCO 2005

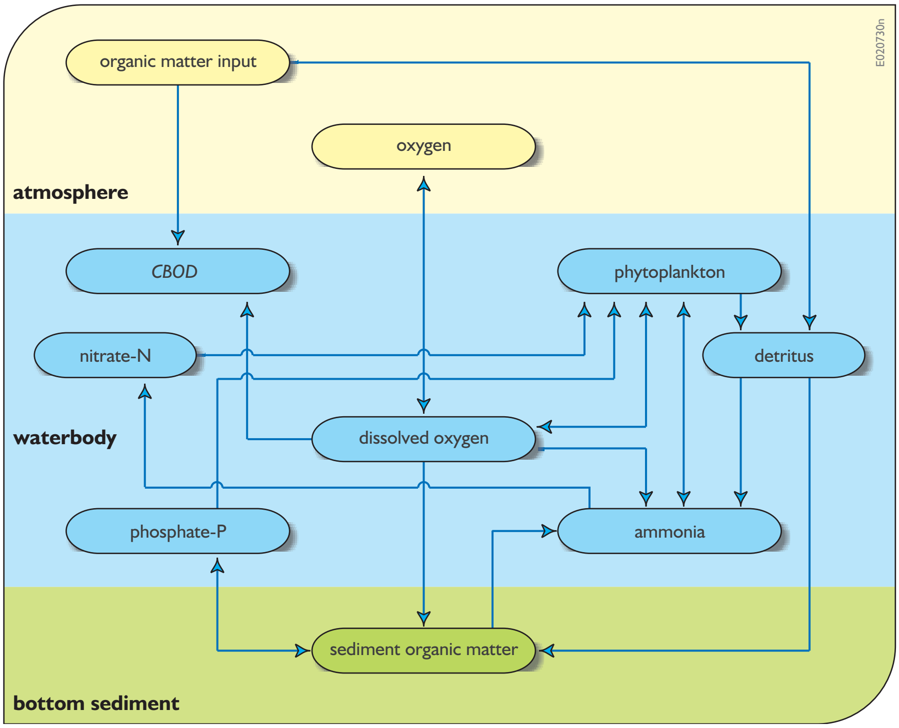



**Figure 12.11.** The dissolved oxygen, nitrogen and phosphorus cycles, and phytoplankton interactions in a water body, showing the decay (satisfaction) of carbonaceous and sediment oxygen demands, re-aeration or deaeration of oxygen at the air–water interface, ammonification of organic nitrogen in the detritus, nitrification (oxidation) of ammonium to nitrate-nitrogen and oxidation of organic phosphorus in the sediment or bottom layer to phosphate phosphorus, phytoplankton production from nitrate and phosphate consumption, and phytoplankton respiration and death contributing to the organic nitrogen and phosphorus.

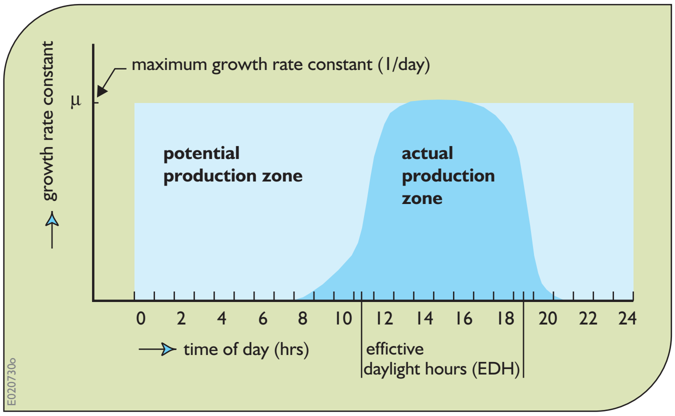



**Figure 12.12.** Calculation of the fraction, $f_d$, of the maximum growth rate constant, $\mu$, to use in the algal growth equations. The fraction $f_d$ is the ratio of actual production zone/potential production zone.

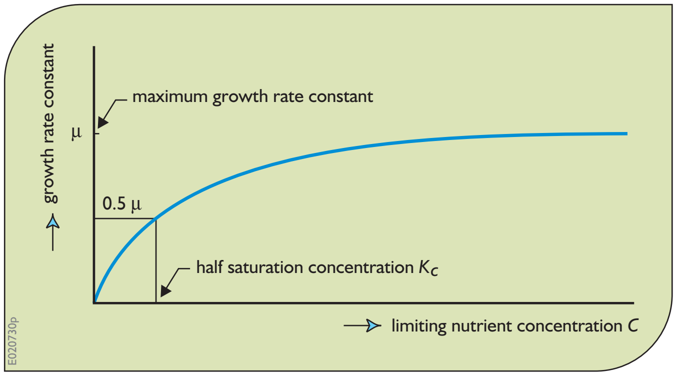



**Figure 12.13.** Defining the half saturation constant for a Michaelis–Menten model of algae. The actual growth rate constant $= \mu C/(C + K_C)$.

Representative values of the coefficients for this model are shown in Table 12.4.

Because the growth of phytoplankton cannot occur without nutrients, the eutrophication modelling must be coupled with that of nutrients. Nutrient modelling must include all the different biochemical forms of the nutrients, primarily nitrogen and phosphorus, as well as all the interactions between the different forms. The sum of all these interactions is referred to a ‘nutrient cycling’.

The nitrogen cycle includes ammonium ($NH_4$-N) and nitrate/nitrite (represented as $NO_3$-N) as the main forms of dissolved nitrogen in water. Furthermore, nitrogen is present in algae, as well as detritus, resulting from algae mortality, and suspended (non-detritus) organic nitrogen. Nitrogen can also be present in different forms in the bottom sediment. Two important reactions in the nitrogen nutrient cycle are nitrification and denitrification, which affect the flux of ammonium and nitrate in the water column.

**Table 12.4.** Typical values of coefficients in the eutrophication model.

E020913e

| parameter | value | | units |
|---|---|---|---|
| $k_N$ half saturation | 10-20 | - a | μg N/l |
| | 50-200 | - c | μg $NO_3$/l |
| | 10 (1-20) | - b | μg N/l |
| $k_P$ half saturation | 1-5 | - a | μg P/l |
| | 20-70 | - c | μg P/l |
| | 10 | - b | μg P/l |
| $a_P$ stoichoimetric ratio | 0.012-0.015 | - c | mg P/mg A |
| $a_N$ stoichoimetric ratio | 0.08-0.09 | - c | mg $NO_3$/mg A |
| μ maximum algae growth rate | 1.5 (1.0-2.0) | - b | 1/day |
| e death algae rate | 0.2-8 | - c | 1/day |
| | 0.1 (0.05-0.025) | - b | 1/day |

a - Thomann and Mueller (1987)
b - Schnoor (1996)
c - Bowie et al. (1985)

Nitrification is the conversion of ammonium to nitrite and finally nitrate, requiring the presence of oxygen:

$$NH_4^+ + 2O_2 \rightarrow NO_3^- + 2H_2O + 2H^+ \tag{12.34}$$

Denitrification is the process occurring during the breakdown (oxidation) of organic matter by which nitrate is transformed to nitrogen gas, which is then usually lost from the water system. Denitrification occurs in anaerobic conditions:

$$NO_3^- \rightarrow N_2 \tag{12.35}$$

The phosphorus cycle is simpler than the nitrogen cycle because there are fewer forms in which phosphorus can be present. There is only one form of dissolved phosphorus, orthophosphorus (also called orthophosphate, $PO_4$-P). Like nitrogen, phosphorus also exists in algae, in detritus and other organic material, as well as in the bottom sediment. Unlike nitrogen, there can also be inorganic phosphorus in the particulate phase.

Further details of the nutrient cycles are given in Section 5.

## 4.9. Toxic Chemicals

Toxic chemicals, also referred to as 'micro-pollutants', are substances that at low concentrations can impair the reproduction and growth of organisms, including fish and human beings. These substances include heavy metals, many synthetic organic compounds (organic micro-pollutants) and radioactive substances.

### 4.9.1. Adsorbed and Dissolved Pollutants

An important characteristic of many of these substances is their affinity with the surface areas of suspended or bottom sediments. Many chemicals preferentially sorb onto surfaces of particulate matter rather than remaining dissolved in water. To model the transport and fate of these substances, the adsorption–desorption process, estimations of the suspended sediment concentration, resuspension from the bottom and settling are required.

Figure 12.14 depicts the adsorption–desorption and first-order decay processes for toxic chemicals and their interaction in water and sediment. This applies to the water and sediment phases in both the water body and the bottom sediments.

The adsorption–desorption model assumes (conveniently but not always precisely) that an equilibrium exists between the dissolved (in water) and adsorbed (on sediments) concentrations of a toxic constituent such as a heavy metal or organic contaminant. This equilibrium follows a linear relationship. The slope of that linear relation is the partition coefficient $K_p$ (litres/kg). This is shown in Figure 12.15.

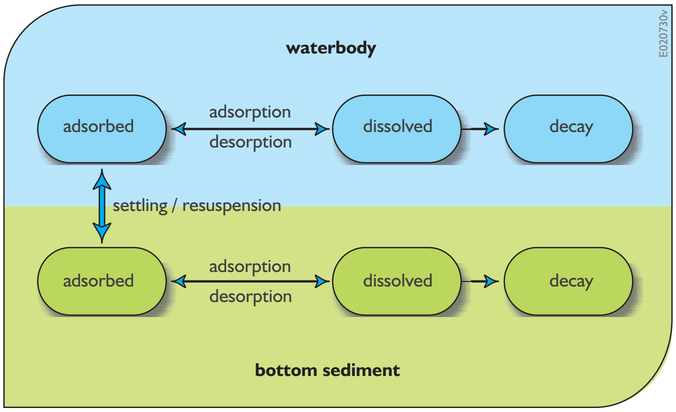



**Figure 12.14.** Schematic of the adsorption/desorption and decay processes of various toxic chemicals in water bodies and bottom sediments.

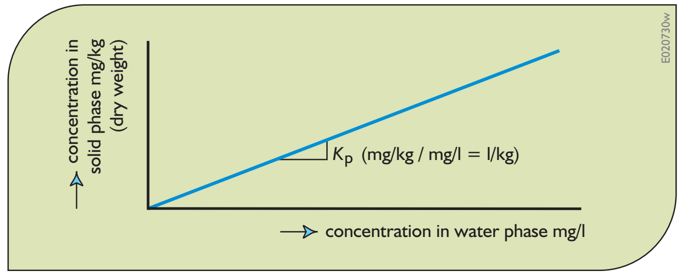



**Figure 12.15.** Defining the partition coefficient $K_p$ (litres/kg) as the slope of the fixed ratio between concentrations of a constituent in the water and sediment phases of either a water body or bottom sediments. Different constituents have different partition coefficients, when they apply.

Each partition coefficient $K_p$ (mg/kgDW/mg/l water or l/kg) is defined as the ratio of the particulate concentration $C'_p$ of a micro-pollutant (mg/kgDW or mg/kgC) divided by the dissolved concentration $C'_d$ of a micro-pollutant (mg/l water).

$$K_p = C'_p / C'_d \tag{12.36}$$

Representative values of partition coefficients $K_p$ are given in Table 12.5.

The presence of a micro-pollutant in a water system is described by the total concentration (sum of dissolved and particulate concentrations), the total particulate concentration and the total dissolved concentration for each water and sediment compartment. The particulate and dissolved concentrations are derived from the total concentration and the respective fractions.

Because the fate of most micro-pollutants is largely determined by adsorption to particulate matter, suspended inorganic and organic matter (including phytoplankton) has to be included in the model in most cases. It may be necessary to include dissolved organic matter as well.

The adsorbed fractions in the water column are subject to settling. The fractions in the sediment are subject to resuspension. The adsorbed fractions in the sediment can also be removed from the modelled part of the water system by burial.

The rates of settling and resuspension of micro-pollutants are proportional to the rates for particulate matter. An additional process called bioturbation leads to

| $K_p$ parameter | value | units |
|---|---|---|
| arsenic | $10^4$ | l/kg |
| heavy metals (Cd,Cu,Cr,Zn) | $10^4$-$10^6$ | l/kg |
| benzo(a)pyrene | $10^4$-$10^5$ | l/kg |
| lead | $10^5$-$10^6$ | l/kg |
| PCB | $10^5$-$10^6$ | l/kg |
| plutonium-239 | $10^4$-$10^5$ | l/kg |
| methoxychlor | $10^4$ | l/kg |
| napthalene | $10^3$ | l/kg |

E020913f

**Table 12.5.** Typical values of partition coefficients in toxic chemical model from Thomann and Mueller (1987).

the redistribution of the micro-pollutant among sediment layers. Bioturbation is caused by the physical activity of organisms, and affects both the particulate and dissolved phases but at different rates. Bioturbation is taken into account by means of dispersion coefficients.

For modelling purposes, it is important to know how much of a toxic chemical is present as a dissolved constituent, as opposed to adsorbed. Assuming partition coefficients apply to a particular toxic constituent, the concentration, $C_w$, of that constituent in the water body is divided into a dissolved fraction ($f_{dw}$) and an adsorbed fraction ($f_{aw}$);

$$C_w = (f_{dw} + f_{aw})C_w \tag{12.37}$$

In turn, the adsorbed fraction is composed the fractions of a micro-pollutant adsorbed to inorganic particulate matter, *fim*, dead particulate organic matter, *fpoc,* and algae, *falg*. The total micro-pollutant concentration, $C_w$ (mg/m$^3$) is the sum of all these fractions:

$$C_w = (f_{dw} + fim + fpoc + falg)C_w \tag{12.38}$$

Considering the simple division into dissolved and adsorbed fractions ($f_{dw}$ and $f_{aw}$), these fractions depend on the partition coefficient, $K_p$, and on the suspended sediment concentration, *SS* (mg/l). The proportions of the total constituent concentration in the water body, $C_w$, dissolved in the water, $DC_w$ (mg/l), and adsorbed to the suspended sediments, $AC_w$ (mg/l) are defined as

$$DC_w = f_{dw}C_w \tag{12.39}$$

$$AC_w = f_{aw}C_w \tag{12.40}$$

where the fractions

$$f_{dw} = 1/(1 + K_pSS) \tag{12.41}$$

$$f_{aw} = 1 - f_{dw} = K_pSS/(1 + K_pSS) \tag{12.42}$$

Similarly, in the bottom sediments, the dissolved concentration $DC_s$ (mg/l) and adsorbed concentration $AC_s$ (mg/l) are fractions, $f_{ds}$ and $f_{as}$, of the total concentration $C_s$ (mg/l);

$$DC_s = f_{ds}C_s \tag{12.43}$$

$$AC_s = f_{as}C_s \tag{12.44}$$

These fractions are dependent on the sediment porosity, $\phi$, and density, $\rho_s$ (kg/l):

$$f_{ds} = 1/[\phi + \rho_s(1 - \phi)K_{\mathrm{p}}] \tag{12.45}$$

$$f_{as} = 1 - f_{ds} = ([\phi + \rho_s(1 - \phi)K_p] - 1)/[\phi + \rho_s(1 - \phi)K_p] \tag{12.46}$$

First order decay occurs in the water and sediment phases only in the dissolved fraction with decay rate constants $k_w$ and $k_s$ (1/day), respectively. Thus:

$$\mathrm{d}C_w/\mathrm{d}t = -k_w\theta_w^{(T-20)}f_{dw}C_w - f_{aw}C_{wS} + f_{as}C_sr \tag{12.47}$$

$$\mathrm{d}C_s/\mathrm{d}t = -k_s\theta_s^{(T-20)}f_{ds}C_s + f_{aw}C_{wS} - f_{as}C_sr \tag{12.48}$$

In the above two equations the parameter *s* represents the mass of settling sediments (mg/day), *r* the mass of resuspension sediments (mg/day), and $\theta$ the temperature correction coefficient of the constituent at temperature $T = 20$ °C. If data are not available to distinguish between the values of the decay rate constants *k* in water and on sediments, they may be assumed to be the same. Similarly, for the values of the temperature correction coefficients $\theta$. The settling and resuspension of suspended solids can be determined each day from a sediment model.

### 4.9.2. Heavy Metals

The behaviour of heavy metals in the environment depends on their inherent chemical properties. Heavy metals can be divided into different categories according to their dissolved form and redox status. Some metals, including copper, cadmium, lead, mercury, nickel, tin and zinc, form free or complexed cations when dissolved in water, e.g. $Cu^{2+}$ or $CuCl^-$. The soluble complexes are formed with negatively charged ions such as chlorine, oxygen or dissolved organic compounds. These heavy metals also tend to form poorly soluble sulphides under chemically reducing conditions. These sulphides generally settle in bottom sediments and are essentially unavailable ecologically. Other metals such as arsenic and vanadium are present as anions in dissolved form. The differences between groups of metals have important consequences for the partitioning of the metals among several dissolved and particulate phases.

Metals are non-decaying substances. The fate of heavy metals in a water system is determined primarily by partitioning to water and particulate matter (including phytoplankton), and by transport. The partitioning divides the total amount of a pollutant into a 'dissolved' fraction and several 'adsorbed' fractions (as described in Equations 12.39–12.42). The fractions of a metal that are adsorbed onto particulate matter are influenced by all the processes that affect particulate matter, such as settling and resuspension.

Partitioning is described in general by sorption to particulates, precipitation in minerals, and complexation in solution. Complexation with inorganic and organic ligands can be considered explicitly in connection with the other processes. Sorption can be modelled as an equilibrium process (equilibrium partitioning) or as the resultant of slow adsorption and desorption reactions (kinetic formulations). In the latter case, partitioning is assumed to proceed at a finite rate proportional to the difference between the actual state and the equilibrium state.

To describe the fate of certain heavy metals in reducing environments, such as sediment layers, the formation of metal sulphides or hydroxides can be modelled. The soluble metal concentration is determined on the basis of the relevant solubility product. The excess metal is stored in a precipitated metal fraction.

Sorption and precipitation affect the dissolved metal concentration in different ways. Both the adsorbed and dissolved fractions increase at increasing total concentration as long as no solubility product is exceeded. When it is, precipitation occurs.

### 4.9.3. Organic Micro-pollutants

Organic micro-pollutants are generally biocides (such as pesticides or herbicides), solvents or combustion products, and include substances such as hexachlorohexane, hexachlorobenzene, PCBs or polychlorobiphenyls, benzo-a-pyrene and fluoranthene (PAHs or polycyclic aromatic hydrocarbons), diuron and linuron, atrazine and simazine, mevinfos and dichlorvos, and dinoseb.

The short-term fate of organic micro-pollutants in a water system is determined primarily by partitioning to water and organic particulate matter (including phytoplankton), and by transport. Additional processes such as volatilization and degradation influence organic micro-pollutant concentrations (this is in contrast to heavy metals, which do not decay). Many toxic organic compounds have decay (or 'daughter') products that are equally, if not more, toxic than the original compound. The rates of these processes are concentration and temperature dependent.

Organic micro-pollutants are generally very poorly soluble in water and prefer to adsorb to particulate matter in the water, especially particulate organic matter and algae. Therefore, the fractions of a micro-pollutant adsorbed to inorganic matter, *fim*, dead particulate organic matter, *fpoc,* the dissolved fraction of a micro-pollutant, *fd,* and algae, *falg*, add up to the total micro-pollutant concentration, $C$ (mg/m$^3$);

$$C = (fd + fim + fpoc + falg)C \tag{12.49}$$

The fractions are functions of the partition coefficients $K_p$ (for algae (m$^3$/g$C$), for inorganic matter (m$^3$/gDW) and for dead particulate organic matter (m$^3$/g$C$)), the individual concentrations $C$ (for algae biomass (g$C$/m$^3$), for inorganic matter (gDW/m$^3$) and for dead particulate organic matter (g$C$/m$^3$)), and the porosity $\phi$ (m$^3$water/m$^3$bulk). In surface water the value for porosity is 1.

$$fd = \phi/[\phi + Kpalg\ Calg + Kpim\ Cim + Kppoc\ Cpoc] \tag{12.50}$$

$$fim = (1 - fd)\ Kpim\ Cim/[Kpalg\ Calg + Kpim\ Cim + Kppoc\ Cpoc] \tag{12.51}$$

$$fpoc = (1 - fd)\ Kppoc\ Cpoc/[Kpalg\ Calg + Kpim\ Cim + Kppoc\ Cpoc] \tag{12.52}$$

$$falg = (1 - fd - fim - fpoc) \tag{12.53}$$

In terms of bulk measures, each partition coefficient $K_p$ (see Equation 12.36) also equals the porosity $\phi$ times the bulk particulate concentration $Cp$ (mg/m$^3$ bulk) divided by the product of the dissolved (mg/l bulk) and particulate (mg/m$^3$ bulk) bulk concentrations, $Cd\ Cs$, all times $10^6$ mg/kg.

$$K_p = 10^6\ \phi Cp/(CdCs) \tag{12.54}$$

Partitioning can be simulated based on the above equilibrium approach or according to slow sorption kinetics. For the latter, the rate, d$Cp$/d$t$, of adsorption or desorption (mg/m$^3$/day) depends on a first order kinetic constant $ksorp$ (day$^{-1}$) for adsorption and desorption times the difference between equilibrium particulate concentration $Cpe$ of a micro-pollutant (mg/m$^3$ bulk) and the actual particulate concentration $Cp$ (mg/m$^3$ bulk) of a micro-pollutant.

$$\mathrm{d}Cp/\mathrm{d}t = ksorp(Cpe - Cp) \tag{12.55}$$

The kinetic constant for sorption is not temperature dependent. All other kinetic constants for micro-pollutants are temperature dependent.

Mass-balance equations are similar for all micro-pollutants except for the loss processes.

Metals are conservative substances that can be transformed into various species either through complexation, adsorption or precipitation. Organic micro-pollutants are lost by volatilization, biodegradation, photolysis, hydrolysis and overall degradation. Most of these processes are usually modelled as first-order processes, with associated rate constants.

Volatilization is formulated according to the double film theory. The volatization rate $dCd/dt$ (mg/m$^3$/day) of dissolved micro-pollutant concentrations, $Cd$ (mg/m$^3$ water), in water depends on an overall transfer coefficient, $kvol$ (m/day), for volatilization and the depth of the water column, $H$ (m);

$$dCd/dt = -\ kvol\ Cd/H \tag{12.56}$$

The numerator ($kvol\ Cd$) is the volatilization mass flux (mg/m$^2$/day).

This equation is only valid when the atmospheric concentration is negligibly small, which is the normal situation.

All other loss rates such as biodegradation, photolysis, hydrolysis or overall degradation (mg/m$^3$/day) are usually modelled as

$$dC/dt = -\ kC \tag{12.57}$$

where $C$ is the total concentration of a micro-pollutant (mg/m$^3$), and $k$ is a (pseudo) temperature-dependent first-order kinetic rate constant for biodegradation, photolysis, hydrolysis or overall degradation (day$^{-1}$).

### 4.9.4. Radioactive Substances

The fate of most radionuclides, such as isotopes of iodine ($^{131}$I) and cesium ($^{137}$Cs), in water is determined primarily by partitioning to water and particulate matter (including phytoplankton), by transport and by decay. Cesium ($Ce^+$) adsorbs to particulate inorganic matter, to dead particulate organic material, and to phytoplankton, both reversibly and irreversibly. The irreversible fraction increases over time as the reversible fraction gradually transforms into the irreversible fraction. Radioactive decay proceeds equally for all fractions. Precipitation of cesium does not occur at low concentrations in natural water systems.

Iodine is only present in soluble form as an anion ($IO_3^-$) and does not adsorb to particulate matter. Consequently, with respect to transport, iodine is only subject to advection and dispersion.

Concentrations of radionuclides, $C_R$ (mg/m$^3$), are essentially conservative in a chemical sense, but they decay by falling apart in other nuclides and various types of radiation. The radioactive decay rate (mg/m$^3$/day) is usually modelled as a first order process involving a kinetic radioactive decay constant, $kdec$ (day$^{-1}$). This kinetic constant is derived from the half-life time of the radionuclide. The initial concentration may be expressed as radioactivity, in order to simulate the activity instead of the concentration. These state variables can be converted into each other using:

$$Ac = 10^{-3}\ N_A kdec C_R/[86400\ Mw] \tag{12.58}$$

where

$Ac$ = activity of the radionuclide (Bq/m$^3$/s)

$N_A$ = Avogadro's number ($6.02 * 10^{23}$ mole)

$Mw$ = molecular weight of the radionuclide (g/mole).

## 4.10. Sediments

As discussed in the previous section, sediments play an important role in the transport and fate of chemical pollutants in water. Natural waters can contain a mixture of particles ranging from gravel (2 mm to 20 mm) or sand (0.07 mm to 2 mm) down to very small particles classified as silt or clay (smaller than 0.07 mm). The very fine fractions can be carried as colloidal suspension, for which electrochemical forces play a predominant role. Considering its large adsorbing capacities, the fine fraction is characterized as cohesive sediment. Cohesive sediment can include silt and clay particles as well as particulate organic matter such as detritus and other forms of organic carbon, diatoms and other algae. Since flocculation and adsorbing capacities are of minor importance for larger particles, they are classified as non-cohesive sediment.

The behaviour of this fine-grained suspended matter affects water quality in several ways. First, turbidity and its effect on the underwater light climate is an important environmental condition for algae growth. The presence of suspended sediment increases the attenuation of light in the water column, which leads to an inhibition of photosynthetic activity, and hence a reduction in primary production. Second, the fate of contaminants in waters is closely related to suspended solids due to their large adsorbing capacities. Like dissolved matter, sediment is transported by advection and by turbulent motion. In addition, the fate of the suspended cohesive sediment is determined by settling and deposition, as well as by bed processes of consolidation, bioturbation and resuspension.

### 4.10.1. Processes and Modelling Assumptions

From a water quality perspective, the three most important sediment processes are sedimentation, resuspension or 'erosion', and 'burial' of sediment.

The modelling of sedimentation and erosion processes originates in part from the Partheniades–Krone concept (Partheniades, 1962; Krone, 1962). In this concept, the bottom shear stress plays an essential role in defining whether or not sedimentation of suspended particles or erosion of bed material will occur. Sedimentation takes place when the bottom shear stress drops below a critical value. Resuspension occurs when the bottom shear stress exceeds a critical value. The calculation of bed shear stress is discussed in Section 4.10.5.

For purposes of modelling chemical concentrations in the bottom sediment, one or more sediment layers are usually defined. Each sediment layer is assumed homogeneous (well mixed). The density of the layer can vary, depending on the variable sediment layer composition. The porosity within a given layer is assumed constant and user-defined. Water quality models do not generally consider horizontal transport of bed sediments. This horizontal transport would result in a change in the amount of sediment present in the bottom, and hence a change in the thickness of the layer of non-cohesive sediment.

### 4.10.2. Sedimentation

A characteristic feature of cohesive sediments is their ability to form aggregates of flocs that settle to the bottom, in a process called sedimentation. Whether a particle will settle to the bottom depends on its size and density, and the chemical conditions of the surrounding water system. Various laboratory and field measurements show that the suspended matter concentration strongly influences the aggregation process and thereby the settling velocities of the aggregates (Krone, 1962). Strong flocs are denser and have larger settling velocities.

Sediment floc aggregation depends on the chemical and physical properties of the sediment, and on salinity and turbulence. At high sediment concentrations (several g/l) the particles hinder each other, decreasing their settling velocity. Turbulence affects the flocculation and therefore the settling velocity in two opposing ways. Increased turbulence will increase the collisions between particles, resulting in larger flocs with higher settling velocities, but increased turbulent shear stresses will break up the flocs and decrease their settling velocities.

Sedimentation occurs when the bottom shear stress drops below a critical value. Total shear stress is the sum of the shear stresses induced by flows and wind waves. Sedimentation rates can depend on salinity concentrations.

For any sediment particle $j$ that settles out of the water column, the rate of settling depends on the flow velocity shear stress, $\tau$ (kg/m/sec$^2$), at the bottom surface–water interface, the critical shear stress, $\tau_j$, for the substance $j$, a zero order sedimentation rate, $ZS_j$ (g/m$^2$/day), and the sedimentation velocity, $VS_j$ (m/day), for the settling substance, $C_j$ (g dry wt./m$^3$/day), all divided by the depth, $H$ (m), of settling in the water column.

If the shear stress $\tau$ at the bottom-water interface is less than the critical shear stress $\tau_j$ for the substance $j$, settling is assumed to take place. The rate of decrease of the substance (g dry wt./m$^3$/day) in the water column due to settling is

$$dC_j/dt = -\,[1 - (\tau/\tau_j)]\,(ZS_j + VS_jC_j)/H \quad \tau \leq \tau_j \tag{12.59}$$

The corresponding flux (g dry wt./m$^2$ of sediment layer area/day) of substance $j$ onto the sediment layer (accretion) is

$$dC_j^{sed}/dt = -\,(dC_j/dt)\,H \tag{12.60}$$

Sedimentation always results in an increase of the substance settling in the upper sediment bed layer. The quantity of sedimentation in one model time-step cannot exceed the available amount of substance in the water column.

No net resuspension is assumed to take place if sedimentation occurs.

### 4.10.3. Resuspension

If the shear stress $\tau$ exceeds the critical shear stress $\tau_j$ for the type substance $j$, then no net settling takes place. Instead, resuspension can take place (also called 'erosion').

Erosion of cohesive bed material occurs when the bed shear forces exceed the resistance of the bed sediment. The resistance of the bed is characterized by a certain critical erosive strength (bottom shear stress). This critical stress is determined by several factors, such as the chemical composition of the bed material, particle size distribution and bioturbation. Erosion of sediment is induced by the bed stress due to flow velocities, tidal and wind-induced advective flows, and surface waves. Erosion is directly proportional to the excess of the applied shear stress over the critical erosive bottom shear stress. One formula for erosion of homogeneous beds is based on Partheniades (1962). The erosion/resuspension flux is limited by the available amount of sediment on the sea bed. Typically a one-layer homogeneous bed is assumed.

The rate of resuspension (g dry wt./$m^2$ of bottom layer area/day) of substance $j$ going into the water column depends on these flow velocity shear stresses, a zero-order resuspension flux $ZR$ (g/$m^2$/day) if any, a resuspension rate constant $kR_j$ ($\text{day}^{-1}$), and the amount of substance $C_j^{\text{sed}}$ in the top sediment layer (g/$m^2$);

$$\mathrm{d}C_j/\mathrm{d}t = [(\tau/\tau_j) - 1](ZR_j + kR_jC_j^{\text{sed}})/H \quad \tau \geq \tau_j \tag{12.61}$$

The corresponding flux (g dry wt./$m^2$ sediment layer area/day) of substance $C_j^{\text{sed}}$ from the sediment layer area (erosion) is

$$\mathrm{d}C_j^{\text{sed}}/\mathrm{d}t = -(\mathrm{d}C_j/\mathrm{d}t)\,H \tag{12.62}$$

No sedimentation is assumed if resuspension occurs. Resuspension results in a decrease of sediment in the upper sediment bed layer. The sedimentation in one model time step cannot exceed the available amount of substance in the sediment layer. The values of $ZR_j$ and $\tau$ strongly depend on the sediment properties and environmental parameter values.

For water quality considerations, the chemical composition of the resuspending sediment is assumed to be the same as that of the bottom sediment. Resuspension of chemical component concentrations, $N_j$ (g/$m^3$/day), in a substance from the sediment layer is simply the substance, $C^{\text{sed}}$ resuspension flux (g dry wt./$m^2$/day) (Equation 12.62) times the fraction, $fr_j$, of component $j$ in the substance divided by the depth, $H$, of water column:

$$\mathrm{d}N_j/\mathrm{d}t = fr_j(\mathrm{d}C^{\text{sed}}/\mathrm{d}t)/H \tag{12.63}$$

### 4.10.4. Burial

'Burial' is a term used to convey that bed sediment is no longer available for resuspension because it has been covered by much newer sediment, i.e. it has been buried. This concept is important especially with respect to contaminated sediments, as it implies that buried contaminated sediments are isolated from the overlying water column and no longer pose a threat to the water quality.

Burial is a sink for sediments that are otherwise susceptible for resuspension. Assuming a fixed active sediment layer thickness, burial occurs when this thickness is exceeded. Establishing this requires knowledge of the depth, $h_j$ (m/g dry wt/$m^2$) of particle size substance $j$ per unit dry wt per square metre. Hence, if the total depth of the active sediment layer, $\sum_j C_j^{\text{sed}} h_j$, is greater than the assumed maximum depth, $D^{\max}$ (m), of active sediment layer, then burial will result. The decrease in available sediment substance $j$ will be in proportion to its contribution, $C_j^{\text{sed}}h_j$, to the total depth, $\sum_j C_j^{\text{sed}}h_j$, of the sediment layer. The burial flux will equal a burial rate constant $k_B$ ($\text{day}^{-1}$ and usually set equal to 1) times the amount $C_j^{\text{sed}}$ of substance $j$ times the depth ratio times the excess depth ratio, $(\sum_j C_j^{\text{sed}}h_j - D^{\max})/(\sum_j C_j^{\text{sed}}h_j)$.

$$\mathrm{d}C_j^{\text{bur}}/\mathrm{d}t = k_B C_j^{\text{sed}}\left[C_j^{\text{sed}}h_j \Big/ \left(\sum_j C_j^{\text{sed}}h_j\right)\right] \times \left[\sum_j C_j^{\text{sed}}h_j - D^{\max}\right] \Big/ \left(\sum_j C_j^{\text{sed}}h_j\right) \tag{12.64}$$

### 4.10.5. Bed Shear Stress

The bed shear stress directly influences the sedimentation and erosion rates. It depends on the flow (currents) and the wind-generated (and sometimes human- generated) surface waves. For sedimentation–erosion processes, it is usually assumed that the total bed shear stress, $\tau$ (Newton/$m^2$ or kg/m/$s^2$) due to flow, $\tau_{\text{flow}}$, and waves, $\tau_{\text{wave}}$, are additive:

$$\tau = \tau_{\text{flow}} + \tau_{\text{wave}} \tag{12.65}$$

The bed shear stress $\tau_{\text{flow}}$ for depth-averaged flow depends on the water density (1,000 kg/$m^3$), the horizontal flow velocity, $U_h$ (m/s), acceleration of gravity, $g$ (9.81 m/$s^2$), and a Chezy coefficient, $Cz$ ($m^{0.5}$/s).

$$\tau_{\text{flow}} = 1{,}000\,(9.81)U_h^2/Cz^2 \tag{12.66}$$

The Chezy coefficient, $Cz$ ($m^{0.5}/s$), is defined as a function of water depth, $H$ (m), and bottom roughness length (Nikuradse equivalent roughness length) (m), *Rough* (m), or Manning's roughness coefficient, n ($s/m^{1/3}$):

$$Cz = 18 \log_{10}(12\ H/Rough) \tag{12.67}$$

$$Cz = (H)^{1/6}/n \tag{12.68}$$

Both Nikuradse's roughness coefficient and Manning's roughness coefficient may be changing due to bed load movements. Here (and in most models) they are assumed fixed.

Surface waves are caused by wind stress on the water surface. The magnitude of the waves depends on the wind conditions, wind duration, water depth and bottom friction. Wave fields are commonly described by the significant wave height, significant wave period and wavelength. Waves induce a vertical circular movement (orbital velocity) that decreases with depth. The waves exert friction forces on the bed during propagation.

The magnitude of the bed shear stress, $\tau_{\text{wave}}$, due to waves depends on a wave friction factor, $f_w$, the density of water, (1,000 $kg/m^3$) and the effective orbital horizontal velocity at the bed surface, $U_o$.

$$\tau_{\text{wave}} = 0.25\ (1{,}000)\ f_w U_o^2 \tag{12.69}$$

The wave friction factor, $f_w$, and the effective orbital horizontal velocity at the bed surface, $U_o$, are functions of three wave parameters: the significant wave height, $H_s$ (m), the mean wave period, $T_m$ (s), and mean wave length, $L_m$ (m) (see also Figure 12.16). Also required is the depth of water, $H$ (m).

The effective horizontal bottom velocity due to waves is defined as

$$U_o = \pi H_s/[T_m \sinh(2\pi H/L_m)] \tag{12.70}$$

The friction or shear factor, $f_w$, can be calculated in two ways (Monbaliu, et al., 1999). One way is

$$f_w = 0.16\ [Rough/(U_o T_m/2\pi)]^{0.5} \tag{12.71}$$

The other way uses a factor depending on a parameter $A$ defined as

$$A = H_s/[2\ Rough \sinh(2\pi H/L_m)] \tag{12.72}$$

If $A > 1.47$ then

$$f_w = \exp\{-5.977 + 5.123H^{-0.194}\} \tag{12.73}$$

Otherwise

$$f_w = 0.32 \tag{12.74}$$

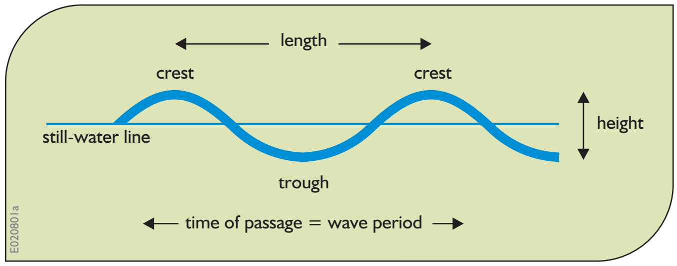



**Figure 12.16. Wave dimensions of significant wave height, period and length.**

## 4.11. Lakes and Reservoirs

The water quality modelling principles discussed above are applicable to all different types of water systems: streams, rivers, lakes, estuaries and even coastal or ocean waters. This section presents some of the unique aspects of water quality modelling in lakes. The physical character and water quality of rivers draining into lakes and reservoirs are governed in part by the velocity and volume of river water. The characteristics of the river water typically undergo significant changes as the water enters the lake or reservoir, primarily because its velocity reduces. Sediment and other material that were carried in the faster-flowing water settle out in the basin.

The structure of the biological communities also changes from organisms suited to living in flowing waters to those that thrive in standing or pooled waters. There are greater opportunities for the growth of algae (phytoplankton) and the development of eutrophication.

Reservoirs typically receive larger inputs of water (as well as soil and other materials carried in rivers) than lakes do, and as a result, may receive larger pollutant loads. However, because of greater water inflows, flushing rates are more rapid than in lakes. Thus, although reservoirs may receive greater pollutant loads than lakes, they have the potential to flush out the pollutants more rapidly. Reservoirs may therefore exhibit fewer or less severe negative water quality or biological impacts than lakes for the same pollutant load.

The water quality of lakes and reservoirs is defined by variables measured within the water basin. Although there are many variables of limnological significance, water quality is typically characterized on the basis of conditions such as:

- water clarity or transparency (greater water clarity usually indicates better water quality)
- concentration of nutrients (lower concentrations indicate better water quality)
- quantity of algae (lower levels indicate better water quality)
- oxygen concentration (higher concentrations are preferred for fisheries)
- concentration of dissolved minerals (lower values indicate better water quality)
- acidity (a neutral pH of 7 is preferred).

Many waste chemical compounds from industry, some with toxic or deleterious effects on humans and/or other water-dependent organisms and products, are discharged into lakes and reservoirs. Some can kill aquatic organisms and damage irrigated crops. Inadequate water purification resulting in the discharge of bacteria, viruses and other organisms into natural waters can be a primary cause of waterborne disease. Although dangerous to human health worldwide, such problems are particularly severe in developing countries.

There can be major differences between deep and shallow lakes or reservoirs. Deep lakes, particularly in non-tropical regions, usually have poorer water quality in their lower layers, due to stratification (see Section 4.11.3). Shallow lakes do not exhibit this depth differentiation in quality. Their more shallow, shoreline areas have relatively poorer water quality because those sites are where pollutant inputs are discharged and have a greater potential for disturbance of bottom muds and the like. The water quality of a natural lake usually improves as one moves from the shoreline to the deeper central part.

In contrast, the deepest end of a reservoir is usually immediately upstream of the dam. Water quality usually improves along the length of a reservoir, from the shallow inflow end to the deeper, 'lake-like' end near the dam, as shown in Figure 12.17.

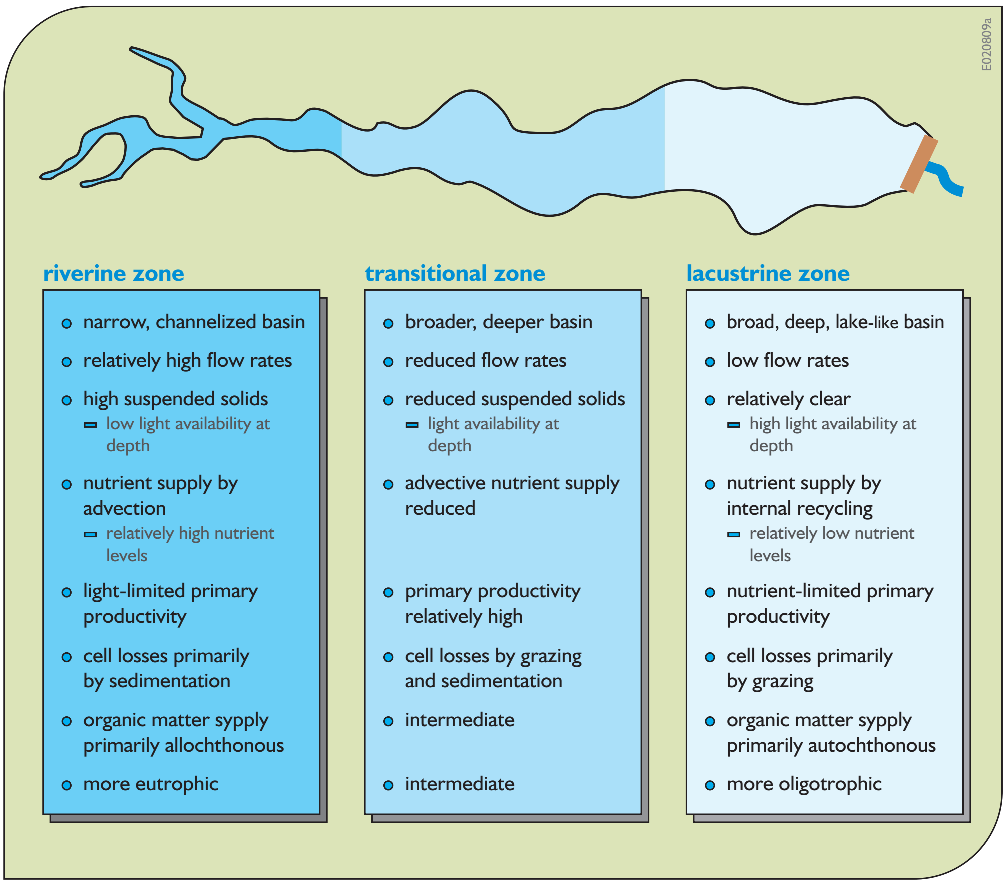



**Figure 12.17.** Longitudinal zonation of water quality and other variables in reservoirs (UNDP, 2000).

Reservoirs, particularly the deeper ones, are also distinguished from lakes by the presence of a longitudinal gradient in physical, chemical and biological water quality characteristics from the upstream river end to the downstream dam end. Because of this, reservoirs have been characterized as comprising three major zones: an upstream riverine zone, a downstream lake-like zone at the dam end, and a transitional zone separating these two (Figure 12.17). The relative size and volume of the three zones can vary greatly in a given reservoir.

### 4.11.1. Downstream Characteristics

Constructing a dam can produce dramatic changes in the downstream river channel below it. These are quite unlike downstream changes from lakes. Because reservoirs act as sediment and nutrient traps, the water at the dam end of a reservoir is typically of higher quality than water entering the reservoir. This higher-quality water subsequently flows into the downstream river channel below the dam. This phenomenon is sometimes a problem, in that the smaller the quantity of sediments and other materials transported in the discharged water, the greater the quantity that it can now pick up and transport. Because it contains less sediment, the discharged water can scour and erode the streambed and banks, picking up new sediment as it continues downstream. This scouring effect can have significant negative impacts on the flora, fauna and biological community structure in the downstream river channel. The removal of sediments from a river by reservoirs also has important biological effects, particularly on floodplains.

Many reservoirs, especially those used for drinking supplies, have water release or discharge structures located at different vertical levels in their dams (Figure 12.18). This allows for the withdrawal or discharge of water from different layers within the reservoir, known as ‘selective withdrawal’. Depending on the quality of the water discharged, selective withdrawal can significantly affect water quality within the reservoir itself, as well as the chemical composition and temperature of the downstream river. Being able to regulate both quantities and qualities of the downstream hydrological regimes makes it possible to affect both flora and fauna, and possibly even the geomorphology of the stream or river.

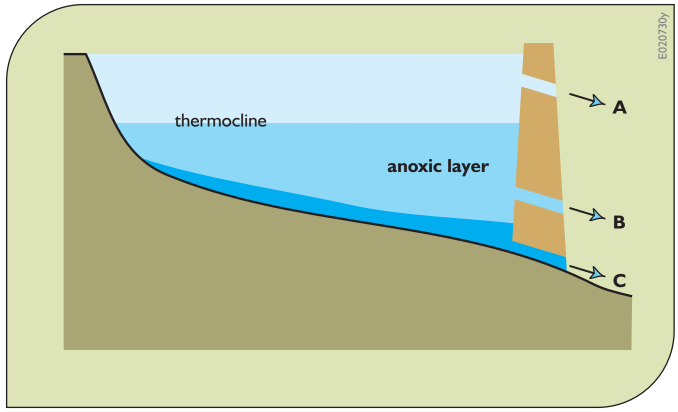



**Figure 12.18.** A multiple-outlet reservoir can be better used to regulate the temperature and water quality downstream.

Constructing a reservoir may have significant social and economic implications, including the potential for stimulating urban and agricultural development adjacent to, and below, the reservoir. These activities can have both positive and negative impacts on downstream water quality, depending on the nature and size of development.

Agriculture is often the leading source of pollution in lakes. Healthy lake ecosystems contain nutrients in small quantities from natural sources, but extra inputs of nutrients (primarily nitrogen and phosphorus) unbalance lake ecosystems. When temperature and light conditions are favourable, excessive nutrients stimulate population explosions of undesirable algae and aquatic weeds. After the algae die, they sink to the lake bottom, where bacteria consume the available dissolved oxygen as they decompose the algae. Fish may die and foul odours may result if dissolved oxygen is depleted.

Heavy metals are another major cause of lake quality impairment. Our knowledge of this is mainly due to the widespread detection of heavy metals in fish tissue samples. Since it is difficult to measure mercury in ambient water, and since it bioaccumulates in fish tissue, fish samples are commonly used to indicate the level of contamination. Common sources of mercury involve atmospheric transport from power-generating facilities and other ‘smoke-stack’ industries.

In addition to nutrient and metal siltation, enrichment by organic wastes that deplete oxygen and by noxious aquatic plants affect lakes and reservoirs. Often, several pollutants and processes affect a single lake. For example, a process such as removal of shoreline vegetation may accelerate erosion of sediment and nutrients into a lake. Extreme acidity (low pH) resulting from acid rain can eliminate fish in isolated lakes. Urban runoff and storm sewers, municipal sewage treatment plants, and hydrological modifications are also sources of lake pollutants.

The prediction of water quality in surface water impoundments is based on mass-balance relationships similar to those used to predict water quality concentrations in streams and estuaries. There are also significant problems in predicting the water quality of lakes or reservoirs compared to those of river and estuarine systems. One is the increased importance of wind-induced mixing processes and thermal stratification. Another, for reservoirs, is the impact of various reservoir-operating policies.

### 4.11.2. Lake Quality Models

Perhaps the simplest way to begin modelling lakes is to consider shallow well-mixed constant-volume lakes subject to a constant pollutant loading. The flux of any constituent concentration, $C$, in the lake equals the mass input of the constituent less the mass output less losses due to decay or sedimentation, if any, all divided by the lake volume $V$ ($m^3$). Given a constant constituent input rate $W_C$ (g/day) of a constituent having a net decay and sedimentation rate constant $k_C$ ($day^{-1}$) into a lake having a volume $V$ ($m^3$) and inflow and outflow rate of $Q$ ($m^3$/day), then the rate of change in the concentration C ($g/m^3$/day) is

$$dC/dt = (1/V)\,(W_C - QC - k_C CV) \tag{12.75}$$

Integrating this equation yields a predictive expression of the concentration $C(t)$ of the constituent at the end of any time period $t$ based in part on what the concentration, $C(t-1)$, was at the end of the previous time period, $t-1$. For a period duration of $\Delta t$ days,

$$\begin{aligned} C(t) = {} & [W_C/(Q + k_C V)] \\ & [1 - \exp\{-\Delta t((Q/V) + k_C)\}] \\ & + C(t-1)\exp\{-\Delta t((Q/V) + k_C)\} \end{aligned} \tag{12.76}$$

The equilibrium concentration, $C_e$, can be obtained by assuming each concentration is equal in Equation 12.76 or by setting the rate in Equation 12.75 to 0, or by setting $C(t-1)$ equal to 0 and letting $\Delta t$ go to infinity in Equation 12.76. The net result is

$$C_e = W_C/(Q + k_C V) \tag{12.77}$$

The time, $t_\alpha$, since the introduction of a mass input $W_C$ that is required to reach a given fraction $\alpha$ of the equilibrium concentration (i.e., $C(t)/C_e = \alpha$) is

$$t_\alpha = -\,V[\ln(1 - \alpha)]/(Q + K_C V) \tag{12.78}$$

Similar equations can be developed to estimate the concentrations and times associated with a decrease in a pollutant concentration. For the perfectly mixed lake having an initial constituent concentration $C(0)$, say after an accidental spill, and no further additions, the change in concentration with respect to time is

$$dC/dt = -\,C\,(Q + k_C V)/V \tag{12.79}$$

Integrating this equation, the concentration $C(t)$ is

$$C(t) = C(0)\exp\{-\,t((Q/V) + K_C) \tag{12.80}$$

In this case, one can solve for the time $t_\alpha$ required for the constituent to reach a fraction $(1 - \alpha)$ of the initial concentration $C(0)$ (that is, $C(t)/C(0) = (1 - \alpha)$). The result is Equation 12.78.

Equation 12.76 can be used to form an optimization model for determining the wasteload inputs to this well-mixed lake that meet water quality standards. Assuming that the total of all natural wasteloads $W_C(t)$, inflows and outflows $Q(t)$, and the maximum allowable constituent concentrations in the lake, $C(t)^{max}$, may vary among different within-year periods $t$, the minimum fraction, $X$, of total waste removal required can be found by solving the following linear optimization model:

$$\text{Minimize } X \tag{12.81}$$

Subject to the following mass-balance and constituent concentration constraints for each period $t$:

$$\begin{aligned} C(t) = {} & [W_C(t)(1-X)/(Q(t) + k_C V)] \\ & [1 - \exp\{-\Delta t((Q(t)/V) + k_C)\}] \\ & + C(t-1)\exp\{-\Delta t((Q(t)/V) + k_C)\} \end{aligned} \tag{12.82}$$

$$C(t) \le C(t)^{max} \tag{12.83}$$

If each period $t$ is a within-year period, and if the wasteloadings and flows in each year are the same, then no initial concentrations need be assumed and a steady state solution can be found. This solution will indicate, for the loadings $W_C(t)$, the fraction $X$ of waste removal that meet the quality standards, $C(t)^{max}$, throughout the year.

### 4.11.3. Stratified Impoundments

Many deep impoundments become stratified during particular times of the year. Vertical temperature gradients arise that imply vertical density gradients. The depth-dependent density gradients in stratified lakes effectively prevent complete vertical mixing. Particularly in the summer season, lakes may exhibit two zones: an upper volume of warm water called the *epilimnion* and a lower colder volume called the *hypolimnion*. The transition zone or boundary between the two zones is called the *thermocline* (Figure 12.18).

Because of lake stratification, many models divide the depth of water into layers, each of which is assumed to be fully mixed. To illustrate this approach without getting into too much detail, consider a simple two-layer lake in the summer that becomes a one-layer lake in the winter. This is illustrated in Figure 12.19

Discharges of a mass $W_C$ of constituent $C$ in a flow $Q^{in}(t)$ into the lake in period $t$ have concentrations of $W_C/Q^{in}(t)$. The concentration in the outflows from the summer epilimnion is $C_e(t)$ for each period $t$ in the summer season. The concentration of the outflows from the winter lake as a whole is $C(t)$ for each period $t$ in the winter season. The summer time rates of change in the epilimnion constituent concentrations $C_e(t)$ and hypolimnion concentrations $C_h(t)$ depend on the mass inflow, $W_C(t)$, and outflow, $C_e(t)Q^{out}(t)$, the net vertical transfer across thermocline, $(v/D_T)[C_h(t)V_h(t) - C_e(t)V_e(t)]$, the settling on sediment interface, $sH_h(t)C_h(t)$, and the decay, $kC_e(t)$:

$$dC_e(t)/dt = (1/V_e(t))\{(W_C(t) - C_e(t)Q^{out}(t)) + (v/D_T)[C_h(t)V_h(t) - C_e(t)V_e(t)]\} - kC_e(t) \quad (12.84)$$

$$dC_h(t)/dt = -kC_h(t) - (v/D_T)[C_h(t) - C_e(t)V_e(t)/V_h(t)] - sH_h(t)C_h(t) \quad (12.85)$$

In the above two equations, $V_e$ and $V_h$ ($m^3$) are the time-dependent volumes of the epilimnion and hypolimnion respectively; $k$ ($day^{-1}$) is the temperature-corrected decay rate constant; $v$ (m/day) is the net vertical exchange velocity that includes effects of vertical dispersion, erosion of hypolimnion, and other processes that transfer materials across the thermocline of thickness $D_T$ (m); $s$ is the settling rate velocity (m/day); and $H_h(t)$ is the average depth of the hypolimnion (m).

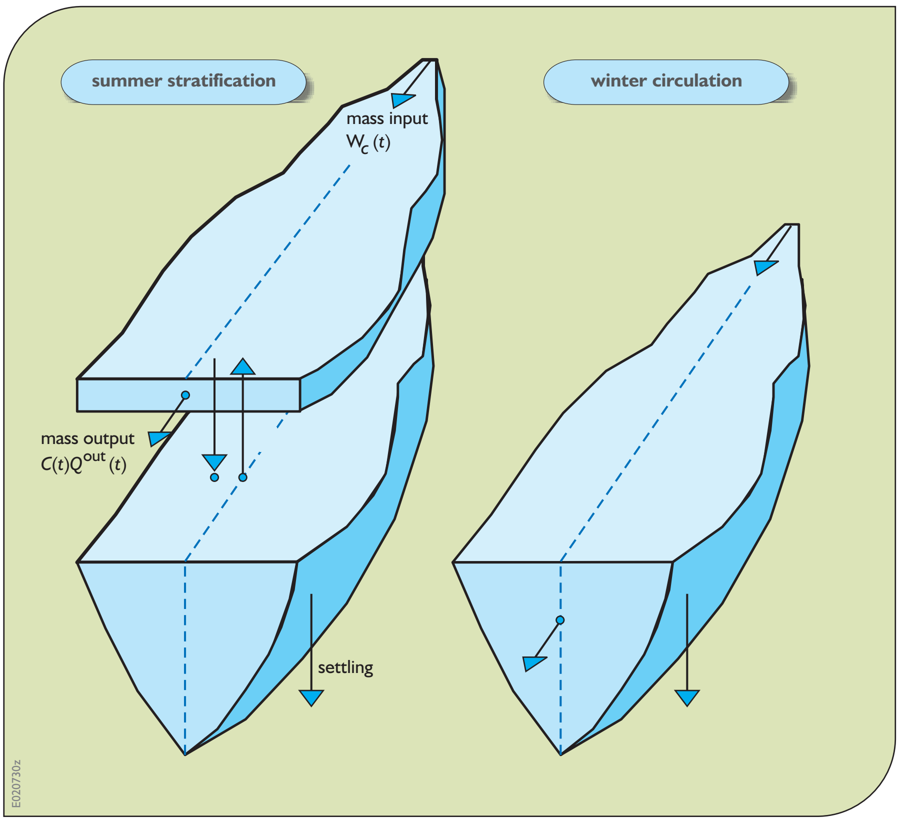



**Figure 12.19.** Lake stratification during summer and complete mixing during winter season.

In the winter season the lake is assumed to be fully mixed. Thus for all periods $t$ in the winter season the initial concentration of a constituent is:

$$C(t) = C_e(t)V_e(t) + C_h(t)V_h(t)/[V_e + V_h] \tag{12.86}$$

$$\mathrm{d}C(t)/\mathrm{d}t = (1/V(t))\{(W_C(t) - C(t)Q^{\text{out}}(t)) - kC_e(t) - sH(t)C(t) \tag{12.87}$$

At the beginning of the summer season, each epilimnion and hypolimnion concentration will be the same:

$$C_e(t) = C(t) \tag{12.88}$$

$$C_h(t) = C(t) \tag{12.89}$$

# 5. An Algal Biomass Prediction Model

An alternative approach to modelling the nutrient, oxygen and algae parts of an ecological model has been implemented in a simulation model developed by WL | Delft Hydraulics called DELWAQ-BLOOM (Los, 1991; Los et al., 1992; Molen et al., 1994; Smits 2001; WL | Delft Hydraulics, 2003). This model is used to predict algae growth and mortality, oxygen concentrations and nutrient dynamics.

The ecological model DELWAQ-BLOOM calculates the advection and dispersion of constituents (state variables) in the water column, and the water quality and ecological processes affecting the concentrations of the constituents. It is based on a three-dimensional version of the governing Equation 12.10. The focus here will be on the source and sink terms in that equation that define the water quality and ecological processes, mostly related to algae growth and mortality, mineralization of organic matter, nutrient uptake and release, and oxygen production and consumption.

For this discussion, consider three nutrient cycles (nitrogen, N, phosphorus, P, and silica, Si) and five different groups of algae, (phytoplankton (diatoms, flagellates and dinoflagellates) or macroalgae ('attached' or 'suspended' *Ulva*)), suspended and bottom detritus, oxygen and inorganic phosphorus particulate matter in the bottom sediments. Different predefined sets of species are available for marine and freshwater ecosystems.

The model processes relating these substances are all interrelated. However, for clarity, the processes can be grouped into nutrient cycling, algae modelling, and oxygen-related processes.

## 5.1. Nutrient Cycling

The DELWAQ-BLOOM model assumes that algae consume ammonia and nitrate in the water column. It includes the uptake of inorganic nutrients by bottom algae, algae-mortality-producing detritus and inorganic nutrients, mineralization of detritus in the water column producing inorganic nutrients, and mineralization of detritus in the bottom that also produces inorganic nutrients. Optionally the model includes algae species with the ability to take up atomic nitrogen (N-fixation) and detritus (mixotrophic growth). The model accounts for the settling of suspended detritus and algae and inorganic adsorbed phosphorus, resuspension of bottom detritus, release of inorganic bottom nutrients to the water, burial of bottom detritus, nitrification or denitrification depending on the dissolved oxygen concentration, and adsorption/desorption of orthophosphate.

## 5.2. Mineralization of Detritus

The oxidation or mineralization of the nutrients in detritus (*DetN*, *DetP*, *DetSi*) and also of detritus carbon (*DetC*) reduces detritus concentrations. The mineralization process consumes oxygen and produces inorganic nutrients ($NH_4$, $PO_4$, and $SiO_2$). The fluxes, $\mathrm{d}C/\mathrm{d}t$, for these four constituents $C$ (mg/l or g/m$^3$) are assumed to be governed by first order processes whose temperature corrected rate constants are $k_C\theta_C^{(T-20)}$ (1/day). Thus:

$$\mathrm{d}C/\mathrm{d}t = k_C\theta_C^{(T-20)}C \tag{12.90}$$

This equation applies in the water column as well as in the bottom sediments, but the mineralization rate constants, $k_C\theta_C^{(T-20)}$, may differ. This rate constant also depends on the stochiometric composition of detritus, relative to the requirements of the bacteria. The concentrations of these detritus constituents in the bottom are usually expressed in grams per square metre of surface area divided by the depth of the sediment layer.

## 5.3. Settling of Detritus and Inorganic Particulate Phosphorus

The rate of settling of nutrients in detritus and inorganic particulate phosphorus out of the water column and on to the bottom is assumed to be proportional to their water column concentrations, $C$. Settling decreases the concentrations of these constituents in the water column.

$$dC/dt = -SR_C(C)/H \qquad (12.91)$$

The parameter $SR_C$ is the settling velocity (m/day) of constituent concentration $C$ and $H$ is the depth (m) of the water column.

## 5.4. Resuspension of Detritus and Inorganic Particulate Phosphorus

The rates at which nutrients in detritus and inorganic particulate phosphorus are resuspended depend on the flow velocities and resulting shear stresses at the bed surface–water column interface. Below a critical shear stress no resuspension occurs. Resuspension increases the masses of these constituents in the water column without changing its volume, and hence increases their concentrations in it. For $C_B$ representing the concentration (grams of dry weight per cubic metre) of resuspended material in the active bottom sediment layer, the flux of constituent concentration in the water column is

$$dC/dt = RR_C C_B/H \qquad (12.92)$$

where $RR_C$ (m/day) is the velocity of resuspension (depending on the flow velocity) and $H$ is the depth of the water column.

## 5.5. The Nitrogen Cycle

The nitrogen cycle considers the water-column components of ammonia ($NH_4$-N), nitrite and nitrate (represented together as $NO_3$-N), algae (*AlgN*), suspended detritus (*DetN*), and suspended (non-detritus) organic nitrogen (*OON*). In the bottom, sediment bottom detritus (*BDetN*) and bottom diatoms (*BDiatN*) are considered. Figure 12.20 shows this nitrogen cycle.

### 5.5.1. Nitrification and Denitrification

Two important reactions in the nitrogen nutrient cycle are nitrification and denitrification. These reactions affect the flux of ammonia and nitrate in the water column. Given sufficient dissolved oxygen and temperature, nitrifying bacteria in the water column transform ammonium to

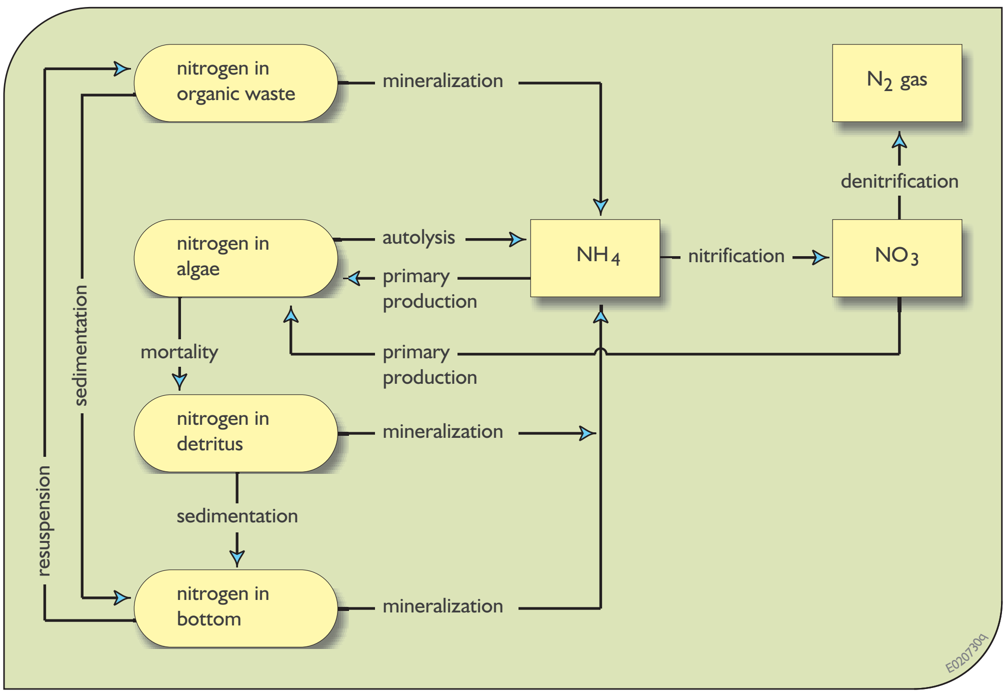



**Figure 12.20.** The nitrogen cycle processes.

nitrite and then nitrate. This can be considered as one reaction,

$$NH_4^+ + 2O_2 \rightarrow NO_3^- + H_2O + 2H^+$$

that occurs at a rate (g*N*/m$^3$/day) of

$$NH_4/dt = -k_{NH_4}\theta_{NH_4}^{(T-20)}NH_4 \tag{12.93}$$

Again, $k_{NH_4}\theta_{NH_4}^{(T-20)}$ is the temperature corrected rate constant (1/day), and $NH_4$ is the concentration of nitrogen in ammonium (gN/m$^3$).

Bacterial activities decrease as temperatures decrease. Bacterial activities also require oxygen. The nitrification process stops if the dissolved oxygen level drops below about 2 mg/l or if the temperature *T* is less than approximately 5 °C.

For each gram of nitrogen in ammonium-nitrogen $NH_4$-N reduced by nitrification, a gram of nitrate-nitrogen $NO_3$-N is produced, consuming 2 moles (64 grams) oxygen per mole (14 grams) of nitrogen (64/14 = 4.57 grams of oxygen per gram of nitrogen). Nitrification occurs only in the water column.

In surface waters with a low dissolved-oxygen content, nitrate can be transformed to free nitrogen by bacterial activity as part of the process of mineralizing organic material. This denitrification process can be written as:

$$\text{'organic matter'} + 2NO_3 \rightarrow N_2 + CO_2 + H_2O \tag{12.94}$$

Nitrate is (directly) removed from the system by means of denitrification. The reaction proceeds at a rate:

$$dNO_3/dt = -k_{NO_3}\theta_{NO_3}^{(T-20)}NO_3 \tag{12.95}$$

where $NO_3$ is the concentration of nitrate nitrogen (gN/m$^3$).

This process can occur both in the water column and the sediment, but in both cases results in a loss of nitrate from the water column. Algae also take up nitrate-nitrogen. As with nitrification, denitrification decreases with temperature. The reaction is assumed to stop below about 5 °C.

### 5.5.2. Inorganic Nitrogen

Ammonia is produced when algae die and the cells release their contents into the surrounding water (in a process called autolysis) and by the mineralization of organic nitrogen in the water and bottom sediment. Ammonia is converted to nitrate by nitrification. Algae use ammonia and nitrate for growth. Different algae prefer either $NH_4$ or $NO_3$ nitrogen. Upon death they release part of their nitrogen contents as ammonia. The remaining nitrogen of dying algae becomes suspended detritus and suspended 'other organic nitrogen' (*OON*). The latter degrades at a much slower rate. Algae can also settle to the bottom. Some macroalgae (*Ulva*) can be fixed to the bottom, unless wind and water velocities are high enough to dislodge them.

Once in the bottom sediment, planktonic algae die and release all their nitrogen contents, as ammonium into the water column and to organic nitrogen in the sediment. In contrast macroalgae and bottom diatoms are attached and are subjected to the normal processes of growth, mortality and respiration.

Suspended detritus and organic nitrogen are formed upon the death of algae. Detritus is also produced by excretion of phyto- and zooplankton and from resuspension of organic matter on and in the sediment. The detritus concentration in the water column decreases by bacterial decay, sedimentation and filtration by zooplankton and benthic suspension feeders.

Bottom detritus is subject to the processes of mineralization, resuspension and burial. Mineralization of bottom detritus is assumed to be slower than that of suspended detritus. The ammonia produced from mineralization is assumed to go directly to ammonia in the water phase. Sedimentation from the water column and mortality of algae in the bottom increase the bottom pool of bottom detritus. The mineralization rate depends on the composition of the detritus, i.e. is a function of the nitrogen/carbon and phosphorus/carbon ratios.

Nitrogen is removed from the system by means of denitrification, a process that occurs under anoxic conditions. Burial is a process that puts the material in a 'deep' sediment layer, and effectively removes it from the active system. This is the only removal process for the other nutrients (P and Si).

## 5.6. Phosphorus Cycle

The phosphorus cycle (Figure 12.21) is a simplified version of the nitrogen cycle. There is only one dissolved pool: orthophosphorus, and only one removal process: burial. However, unlike nitrogen and silica, there is also inorganic phosphorus in the particulate phase (*AAP*).

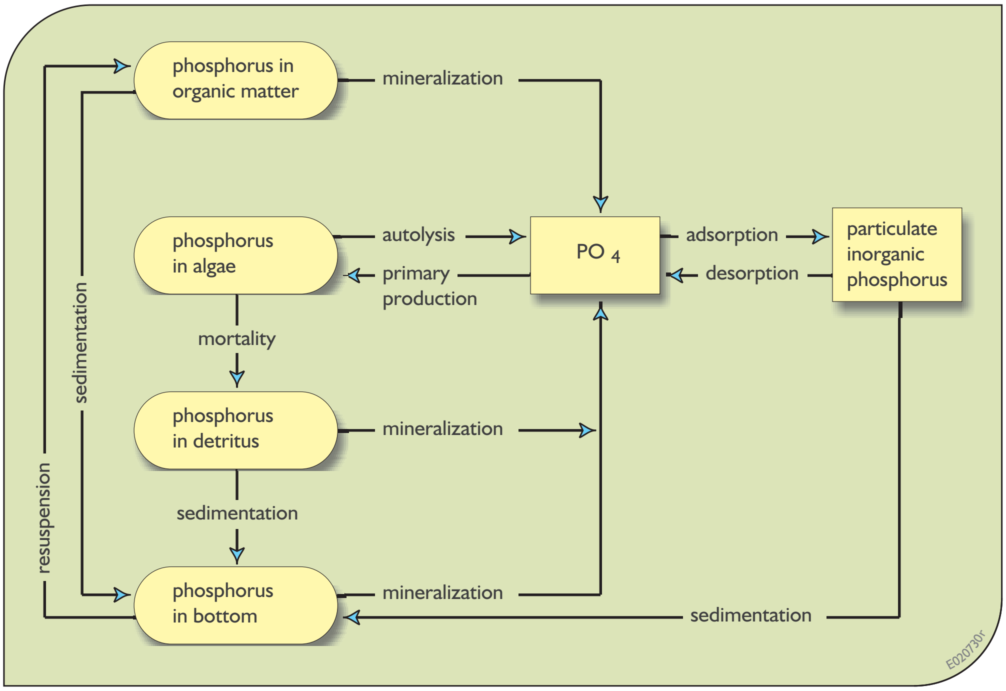



**Figure 12.21.** The processes involved in the phosphorus cycle.

The phosphorus cycle in the water column includes orthophosphate ($PO_4$), algae (*AlgP*), suspended detritus (*DetP*), suspended (non-detritus) organic phosphorus (*OOP*), inorganic adsorbed (available) *P* (*AAP*), and inorganic adsorbed (unavailable) *P* (*UAP*). In the bottom sediment the cycle includes the bottom detritus (*BDetP*) and the bottom inorganic adsorbed *P* (*BAAP*).

A reaction specific to the phosphorus cycle is the adsorbtion/desorption of particulate inorganic phosphorus. Inorganic phosphorus can be present in the aquatic environment in a dissolved form and adsorbed to inorganic particles, such as calcium or iron. The transition from one form into another is not a first order kinetic process, yet in many models desorption of inorganic phosphorus is assumed to be such.

## 5.7. Silica Cycle

The silica cycle is similar to the phosphorus cycle except that there is no adsorption of silica to inorganic suspended solids. Silica is only used by diatoms, so uptake by algae depends on the presence of diatoms. The silica cycle is shown in Figure 12.22.

The silica cycle in the water column includes dissolved silica (Si), diatoms (*Diat*), suspended detritus (*DetSi*), and suspended (non-detritus) organic silica (*OOSi*). In the bottom sediment the cycle includes the bottom detritus (*BDetSi*).

## 5.8. Summary of Nutrient Cycles

The nutrient cycles just described are based on the assumption that nutrients can be recycled an infinite number of times without any losses other than due to transport, chemical adsorption, denitrification and burial. This is an over-simplification of the organic part of the

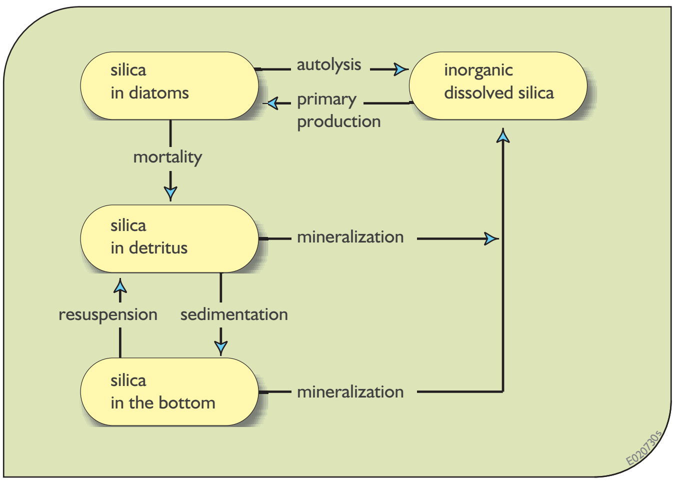



**Figure 12.22.** The processes involved in the silica cycle.

nutrient cycles. The elementary composition of living algae cells is a complicated function of their characteristics as well as the environmental conditions. Upon dying, the algae cell contents are released into the surrounding water. As the cell breaks apart, many of the nutrients are in a form that makes them instantly available for algae cell growth (autolysis). The remaining material consists of more or less degradable substances. Most of this material is mineralized either in the water or at the bottom, but a small portion degrades very slowly if at all. Most of this material settles and is ultimately buried. Resuspension delays but does not stop this process by which nutrients are permanently removed from the water system.

For simplicity, all possible removal processes are lumped into a single term, which is modelled as burial. For example, if a nominal value of 0.0025 per day is used, this means that 0.25% of the bottom amount is buried each day.

The same formulation is used for all three nutrients. Whether or not this is correct depends on the actual removal process. If deactivation is mainly burial into deeper layers of the sediment, there is no reason to distinguish between different nutrients. Other processes such as chemical binding, however, may deactivate phosphorus, but not nitrogen or silica.

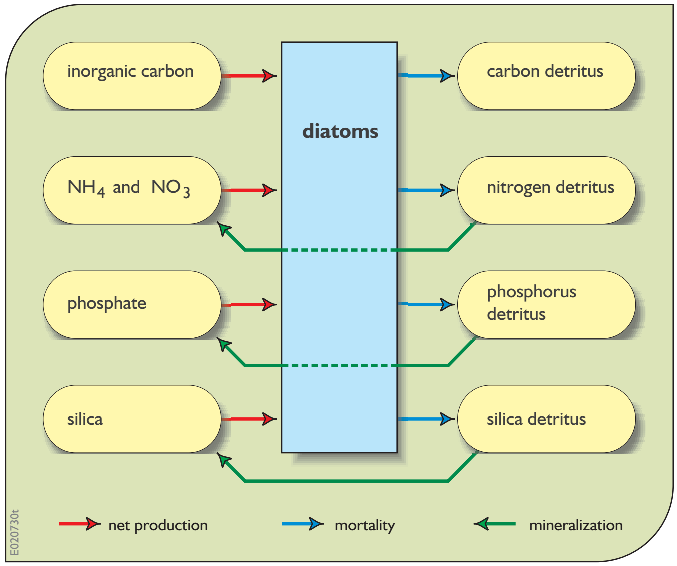



**Figure 12.23. Modelling of diatoms.**

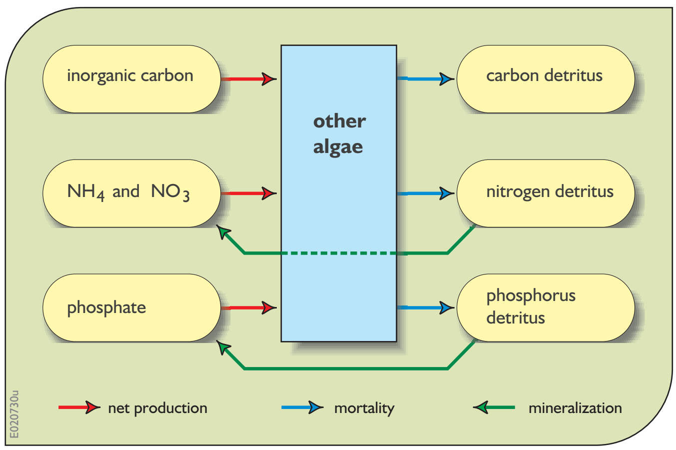



**Figure 12.24. Modelling of other algae besides diatoms.**

## 5.9. Algae Modelling

Algal processes include primary production, mortality (producing detritus and inorganic nutrients) and grazing, settling to become bottom algae and the resuspension of bottom algae, the mortality of bottom algae to bottom detritus, and the burial of bottom algae.

The basic behaviour of algae in surface water can be illustrated by the two diagrams in Figures 12.23 and 12.24. These show the nutrient and carbon fluxes for diatoms and other algae. Diatoms are distinguished from other algae in that they need silicate to grow.

### 5.9.1. Algae Species Concentrations

The module BLOOM II computes phytoplankton within DELWAQ-BLOOM. It is based upon the principle of competition between different species. The basic variables of this module are called 'types'. A type represents the physiological state of a species. Usually a distinction is made between three different types: an *N*-type for nitrogen limitation, a *P*-type for phosphorus limitation and an *E*-type for light energy limitation, whose nutrient content is high (luxury uptake). The solution algorithm of the model considers all potentially limiting factors and first selects the one that is most likely to become limiting. It then selects the best-adapted type for that limitation. The suitability of a type (its fitness) is determined by the ratio of its requirement for that particular resource and its growth rate in equal proportion. This means that a type can become dominant either because it needs a comparatively small amount of a limiting resource (it is efficient) or because it grows rapidly (it is opportunistic).

Now the algorithm considers the next potentially limiting factor and again selects the best-adapted phytoplankton type. This procedure is repeated until it is impossible to select a new pair of a type and limiting factor without violating (that is, over-exhausting) some limiting factor. Thus the model seeks the optimum solution consisting of $n$ types and $n$ limiting factors. To that purpose, it uses linear programming. Typically, BLOOM II considers between three and ten representative algae species and six to fifteen types.

As a further refinement BLOOM II takes the existing biomasses of all phytoplankton types into account. These are the result of growth, loss processes and transport during the previous time steps. Thus, the optimization algorithm does not start from scratch.

As they represent different physiological stages of the same species, the transition of one type to another is a rapid process with a characteristic time step in the order of a day. The equations of the model are formulated in a way that allows such a rapid shift between types of the same species. A transition between different species is a much slower process as it depends on mortality and net growth rates.

It is interesting that the principle just described, by which each phytoplankton type maximizes its own benefit, effectively means that the total net production of the phytoplankton community is maximized.

Denote each distinct species subtype (from now on called type) by the index $k$. The BLOOM II module identifies the optimum concentration of biomass, $B_k$, of each algae type $k$ that can be supported in the aquatic environment characterized by light conditions and nutrient concentrations. Using the net growth rate constant $Pn_k$, it will

$$\text{Maximize } \sum_k Pn_k B_k \tag{12.96}$$

The sum of the biomass concentrations over all types $k$ is the total algae biomass concentration.

For each algae type, the requirements for nitrogen, phosphorus and silica (used only by diatoms) are specified by coefficients $n_{ik}$, the fraction of nutrient $i$ per unit biomass concentration of algae type $k$.

The total readily available concentration, $C_i$ (g/m$^3$) of each nutrient in the water column equals the amount in the total living biomass of algae, $\sum_k(n_{ik}B_k)$, plus the amount incorporated in dead algae, $d_i$, plus that dissolved in the water, $w_i$. These mass-balance constraints apply for each nutrient $i$.

$$\sum_k (n_{ik}\, B_k) + d_i + w_i = C_i \tag{12.97}$$

The unknown concentration variables $B_k$, $d_i$ and $w_i$ are non-negative. All nutrient concentrations $C_i$ are the modelled (optionally measured) total concentrations and are assumed to remain constant throughout the time period (typically a day).

### 5.9.2. Nutrient Recycling

A certain amount of each algae type $k$ dies in each time step, and this takes nutrients out of the live phytoplankton pool. Most of it remains in the detritus and other organic nutrient pools; a smaller fraction (in the order of 30%) is directly available to grow new algae because the dead cells break apart (autolysis) and are dissolved in the water column. Detritus may be removed to the bottom or to the dissolved nutrient pools at rates in proportion to its concentration. Needed to model this are the mortality rate, $M_k$ (day$^{-1}$), of algae type $k$; the fraction, $f_p$; of dead phytoplankton cells that is not immediately released when a cell dies; the remineralization rate constant, $m_i$ (day$^{-1}$), of dead phytoplankton cells; the fraction, $n_{ik}$, of nutrient $i$ per unit biomass concentration of algae type $k$; and the settling rate constant, $s$ (day$^{-1}$), of dead phytoplankton cells.

The rate of change in the nutrient concentration of the dead phytoplankton cells, $\mathrm{d}d_i/\mathrm{d}t$, in the water column equals the increase due to mortality less that which remineralizes and that which settles to the bottom;

$$\mathrm{d}d_i/\mathrm{d}t = \sum_k (f_p M_k n_{ik} B_k) - m_i d_i - s d_i \tag{12.98}$$

Both mortality and mineralization rate constants are temperature dependent.

### 5.9.3. Energy Limitation

Algae absorb light for photosynthesis. Energy becomes limiting through self-shading when the total light absorption of algae, called light extinction, exceeds the maximum at which primary production is just balanced by respiration and mortality. For each algae type $k$ there exists a specific extinction value $K_k^{max}$ (1/m) at which this is the case. The light intensity can also be too high, which means the total extinction is too low (photo-inhibition) for growth. This specific extinction value is $K_k^{min}$. The ranges between $K_k^{min}$ and $K_k^{max}$ differ for different algal types $k$. Letting $K_k$ (m$^3$/m/gdry) represent the specific light absorbing extinction constant for living material of algae type $k$, the total extinction due to all living algae is

$$KL = \sum_k (K_k B_k) \tag{12.99}$$

Added to this must be the extinction caused by dead cells, *KD,* and the background extinction of the water, *KW* (1/m). As defined here, *KW* includes the extinction by inorganic suspended matter and humic substances.

$$K_k^{min} \leq KL + KD + KW \leq K_k^{max} \tag{12.100}$$

In an alternate formulation used in DELWAQ-BLOOM, the total light extinction coefficient in the water column (in m$^{-1}$) is the sum of background extinction ($K_b$, water and non-modelled substances) and extinction due to detritus ($K_{det}$), inorganic material ($K_{im}$), dissolved organic matter or 'yellow substances' ($K_{dom}$) and phytoplankton ($K_{phyt}$):

$$\text{Extinction} = K_b + K_{det} + K_{im} + K_{dom} + K_{phyt} \tag{12.101}$$

The extinction from each component is calculated by its concentration multiplied by its specific extinction coefficient.

The ratio between the various extinction components is highly site specific. In general it also varies seasonally. The amount of dead cells not yet mineralized is, from Equation 12.98, $\sum_k(fpM_kB_k)$. Assuming some fraction $e_d$ (usually between 0.2 and 0.4) of the extinction rate of live cells,

$$KD = e_d \sum_k K_k fpM_k B_k \tag{12.102}$$

If the total extinction is not within the range for an algae type $k$, then its concentration $B_k$ will be zero. To ensure that $B_k$ is 0 if the total extinction is outside of its extinction range, a 0, 1 binary (integer) unknown variable $Z_k$ is needed for each algae type $k$. If $Z_k$ is 1, $B_k$ can be any non-negative value; if it is 0, $B_k$ will be 0. This is modelled by adding three linear constraints for each algae type $k$.

$$KL + KD + KW \leq K_k^{max} + KM\,(1 - Z_k) \tag{12.103}$$

$$KL + KD + KW \geq K_k^{min}\,(Z_k) \tag{12.104}$$

$$B_k \leq BMZ_k \tag{12.105}$$

where *KM* and *BM* are any large numbers no less than the largest possible value of the total extinction or biomass concentration, respectively. Since the objective of maximizing the sum of all $B_k$ together with Equation 12.105 is to set each $Z_k$ value equal to 1, it is only when the total extinction is outside of the extinction range $K_k^{min}$ to $K_k^{max}$ that the $Z_k$ value will be forced to 0. Equation 12.105 then forces the corresponding $B_k$ to 0.

### 5.9.4. Growth Limits

For all algae types $k$, the maximum potential biomass concentration, $B_k^{max}$ (g dry/m$^3$), at the end of the time interval $\Delta t$ (days) depends on the initial biomass concentration, $B_k$, (g dry/m$^3$); the maximum gross production rate, $Pg_k^{max}$ (day$^{-1}$); the respiration rate constant, $R_k$ (day$^{-1}$); and a time and depth averaged production efficiency factor, $E_k$. Mortality is not included in this computation. Using the net production rate constant, $Pn_k$ $(= Pg_k^{max} E_k - R_k)$ (day$^{-1}$), for each algae type $k$:

$$B_k^{max} = B_k^0 \exp\{Pn_k \Delta t\} \tag{12.106}$$

This condition is taken into account by the optimization algorithm. For types with an initial biomass of zero, a small base level is used instead in order to allow previously absent species to start growing.

### 5.9.5. Mortality Limits

As in the case of growth, the mortality of each algae species is also constrained to prevent a complete removal within a single time step. The minimum biomass value of a species is obtained when there is no production, but only mortality. The minimum biomass, $B_k^{min}$ (g dry/m$^3$), of type $k$ at the end of time interval $\Delta t$ depends on the initial biomass, $B_k^0$ (g dry/m$^3$), of type $k$ and the specific mortality rate constant, $M_k$ (day$^{-1}$), of type $k$.

$$B_k^{min} = B_k^0 Z_k \exp\{-M_k \Delta t\} \tag{12.107}$$

These maximum and minimum values are computed for each individual algae type. However, the model sums each of these maximum and minimum values over all subtypes within each species and applies it to the total biomass of the species. This way the maximum possible mortality cannot be exceeded, but transitions between limit types remain possible.

$$\sum_{k \text{ of species } j} B_k^{\min} \leq \sum_{k \text{ of species } j} B_k \leq \sum_{k \text{ of species } j} B_k^{\max} \quad \forall \text{ species } j \tag{12.108}$$

As mortality is computed according to a negative exponential function, the minimum biomass level is always positive; in other words, a species can never disappear completely. For computational purposes, the minimum biomass of a species is set to zero if it drops below some small threshold base level. It is conceivable that the amount of biomass, which should minimally be maintained according to Equation 12.108 exceeds the amount permitted according to the available amount of light energy (12.100). If this happens, Condition (12.100) is dropped.

### 5.9.6. Oxygen-Related Processes

The oxygen concentration in the water column depends in part on the biochemical and physical processes that either produce or consume oxygen. For the algae the model includes the production of oxygen (primary production), and its consumption through respiration. Oxygen is also consumed by the mineralization of detritus and other organic material (in the water column and bottom sediment), nitrification of ammonia to nitrate, and exchange of oxygen with the atmosphere (e.g. re-aeration).

The mineralization of carbon detritus in the water column and bottom consumes oxygen at a molar ratio of 1:1, equivalent to a ratio of 32/12 g $O_2$ to 1 g C. The net growth of algae produces oxygen at a molar ratio of 1:1, equivalent to a ratio of 32/12 g $O_2$ to 1 g C. The mineralization of organic carbon in waste and carbon detritus in the water column and bottom consumes oxygen at a molar ratio of 1:1, equivalent to a ratio of 32/12 g $O_2$ to 1 g C. The nitrification reaction consumes oxygen at a molar ratio of 2:1, equivalent to a ratio of 64/14 g $O_2$ to 1 g N. For all algae, oxygen is produced during photosynthesis. The net growth of algae produces oxygen at a molar ratio of 1:1, equivalent to a ratio of 32/12 g $O_2$ to 1 g C.

#### Bottom Oxygen

The mineralization of organic material in the bottom sediment consumes oxygen, which must be supplied from the water column. The process consumes oxygen at a molar ratio (oxygen to carbon) of 1:1, equivalent to a ratio of 32/12 g $O_2$ to 1 g C.

#### Daily Oxygen Cycle

Because oxygen is produced by the photosynthesis of algae during the daylight hours, there is a natural variation of oxygen concentrations over the twenty-four-hour day–night cycle. Typically, oxygen concentrations are lowest in the early morning as oxygen is consumed during the night through the processes of algae respiration and organic material mineralization. During the daylight hours, oxygen is produced, and the highest values (often supersaturated) are typically found in the late afternoon. When biomasses are high, these variations may be large enough to cause low oxygen conditions during the night or in the early morning.

In the traditional BLOOM calculation, the water quality processes are all calculated for a daily averaged situation. This is reflected by the choice of the input parameters for the light model: the daily averaged solar radiation and the day length. Reducing the time step would be the most straightforward way to include diurnal variations. The drawback, however, is a considerable increase in computation time. Thus an alternative approach has been adopted in the model. The total daily rate of primary production is computed first. Next, this production is distributed over the day. The model takes into account the day length, and oxygen production begins in the first daylight hour. Oxygen production increases during the morning, levels off at a (user-defined) maximum value for a period in the middle of the day, and decreases during the afternoon. There is no oxygen production during the night-time. The hourly oxygen production is combined with the daily averaged oxygen consumption processes and the re-aeration to produce an hourly value of oxygen concentration in the water.

#### Maintenance Respiration

Respiration in algae is a process in which organic carbon is oxidized, using oxygen to produce energy. The process occurs throughout the day and results in oxygen consumption at a rate of 32 grams of oxygen per 12 grams of carbon: 32/12 g $O_2$ to 1 g C. Total respiration is divided into growth and maintenance respiration. Maintenance respiration is a first-order temperature dependent process. Species growth respiration is accounted for when calculating the net primary production. Maintenance respiration, and thus the amount of oxygen consumed, is governed by the temperature-dependent respiration rate constant, $k_k^{\text{res}}\theta_k^{\text{res}(T-20)}$ ($\text{day}^{-1}$), the temperature, $T$ (°C), of the water (higher respiration at higher temperatures), and the concentration of algae biomass, $AlgC_k$. Each algae type can have a different respiration rate. Hence, the rate ($gC_2/m'/day$) at which carbon in algae is oxidized is

$$dAlgC_k/dt = k_k^{\text{res}}\theta_k^{\text{res}(T-20)}AlgC_k \tag{12.109}$$

## 6. Simulation Methods

Most of those who will be using water quality models will be using simulation models that are commonly available from governmental agencies (e.g. USEPA), universities, or private consulting and research institutions such as the Danish Hydraulics Institute, Wallingford software or WL | Delft Hydraulics (Ambrose et al., 1996; Brown and Barnwell, 1987; Cerco and Cole, 1995; DeMarchi et al., 1999; Ivanov et al., 1996; Reichert, 1994; USEPA, 2001; WL | Delft Hydraulics, 2003).

These simulation models are typically based on numerical methods that incorporate a combination of plug flow and continuously stirred reactor approaches to pollutant transport. Users must divide streams, rivers, and lakes and reservoirs into a series of well-mixed segments or volume elements. A hydrological or hydrodynamic model calculates the flow of water between all of these. In each simulation time step, plug flow enters these segments or volume elements from upstream segments or elements. Flow also exits from them to downstream segments or elements. During this time the constituents can decay or grow, as appropriate, depending on the conditions in those segments or volume elements. At the end of each time-step, the volumes and their constituents within each segment or element are fully mixed. The length of each segment or the volume in each element reflects the extent of dispersion in the system.

### 6.1. Numerical Accuracy

Water quality simulation models based on physical, biological and chemical processes typically include time rate of change terms such as $dC/dt$. While it is possible to solve some of these differential equations analytically, most water quality simulation models use numerical methods. The purpose of this section is not to explain how this can be done, but rather to point to some of the restrictions placed on the modeller because of these numerical methods.

Consider first the relationship between the stream, river or lake segments and the duration of time steps, $\Delta t$. The basic first-order decay flux, $dC/dt$ ($g/m^3/day$), for a constituent concentration, $C$, that is dependent on a rate constant, $k$ ($\text{day}^{-1}$), is:

$$dC/dt = -kC \tag{12.110}$$

The finite difference approximation of this equation can be written

$$C(t+\Delta t) - C(t) = -C(t)k\Delta t \tag{12.111}$$

or

$$C(t+\Delta t) = C(t)(1-k\Delta t) \tag{12.112}$$

This equation can be used to illustrate the restriction placed on the term $k\Delta t$. That term cannot exceed a value of 1 or else $C(t+\Delta t)$ will be negative.

Figure 12.25 is a plot of various values of $C(t+\Delta t)/C(t)$ versus $k\Delta t$. This plot is compared with the analytical solution resulting from the integration of Equation 12.110, namely:

$$C(t+\Delta t) = C(t)\exp\{-k\Delta t\} \tag{12.113}$$

Reducing the value of $\Delta t$ will increase the accuracy of the numerical solution. Hence, for whatever value of $\Delta t$, it can be divided by a positive integer $n$ to become $1/n^{\text{th}}$ of its original value. In this case the predicted concentration $C(t+\Delta t)$ will equal

$$C(t+\Delta t) = C(t)(1-k\Delta t/n)^n \tag{12.114}$$

For example if $k\Delta t = 1$, and $n = 2$, the final concentration ratio will equal

$$C(t+\Delta t)/C(t) = (1-1/2)^2 = 0.25 \tag{12.115}$$

Compare this 0.25 to 0.37, the exact solution, and to 0.0, the approximate solution when $n$ is 1. Having $n = 2$ brings a big improvement. If $n = 3$, the concentration

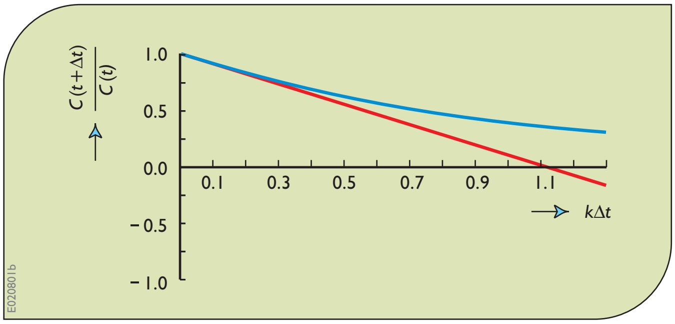



**Figure 12.25.** Plot of numerical approximation (red line) based on Equation 12.112 compared to the true analytical (blue line) value obtained from Equation 12.113.

ratio will be 0.30, an even greater improvement compared to 0. However, no matter what value of $n$ is selected, the predicted concentration will always be less than the actual value based on Equation 12.113, and hence the error is cumulative. Whenever $\Delta t > n/k$, the predicted concentrations will alternate between positive and negative values, either diverging, converging or just repeating the cycle, depending on how much $\Delta t$ exceeds $n/k$. In any event, the predicted concentrations are not very useful.

Letting $m = -n/k\Delta t$, Equation 12.114 can be written as

$$C(t + \Delta t) = C(t)(1 + 1/m)^{m\,(-k\Delta t)} \qquad (12.116)$$

As $n$ approaches infinity, so does the variable m, and hence the expression $(1 + 1/m)^m$ becomes the natural logarithm base $e = 2.718282$. Thus, as $n$ approaches infinity, Equation 12.114 becomes Equation 12.113, the exact solution to Equation 12.110.

## 6.2. Traditional Approach

Most water quality simulation models simulate quality over a consecutive series of discrete time periods. Time is divided into discrete intervals and the flows are assumed constant within each of those time period intervals. Each water body is divided into segments or volume elements, and these are considered to be in steady-state conditions within each simulation time period. Advection or plug flow (i.e. no mixing or dispersion) is assumed during each time period. At the end of each period mixing occurs within each segment or volume element to obtain the concentrations in the segment or volume element at the beginning of the next time step.

This method is illustrated in Figure 12.26. The indices $i-1$, $i$ and $i+1$ refer to stream or river reach segments. The indices $t$ and $t+1$ refer to two successive time periods, respectively. At the beginning of time period $t$, each segment is completely mixed. During the time interval $\Delta t$ of period $t$, the water quality model predicts the concentrations, assuming plug flow in the direction of flow from segment $i$ toward segment $i+1$. The time interval $\Delta t$ is such that the flow from any segment $i$ does not pass through any following segment $i+1$. Hence, at the end of each time period each segment has some of the original water that was there at the beginning of the period, and its end-of-period concentrations of constituents, plus some of the immediately upstream segment's water and its end-of-period concentrations of constituents. These two volumes of water and their respective constituent concentrations are then mixed to achieve a constant concentration within the entire segment. This is done for all segments in each time step. Included in this plug flow and then mixing process are the inputs to the reach from point and non-point sources of constituents.

In Figure 12.26 a mass of waste enters reach $i$ at a rate of $W_i^t$. The volume in each reach segment is denoted by $V$, and the flows from one segment to the next are denoted by $Q$. The drawing on the left represents a portion of a stream or river divided into well-mixed segments. During a period $t$, waste constituents enter segment $i$ from the immediate upstream reach $i-1$ and from the point waste source. In this illustration, the mass of each of these wastes is assumed to decay during each time period, independent of other wastes in the water. Depending on the types of waste, the decay (or even growth) processes that take place may be more complex than those assumed in this illustration. At the end of each time period these altered wastes are mixed together to create an average concentration for the entire reach segment. This illustration applies for each reach segment $i$ and for each time period $t$.

The length, $\Delta x_i$, of each completely mixed segment or volume element depends on the extent of dispersion. Reducing the length of each reach segment or size of each volume element reduces the dispersion within the entire stream or river. Reducing segment lengths, together with

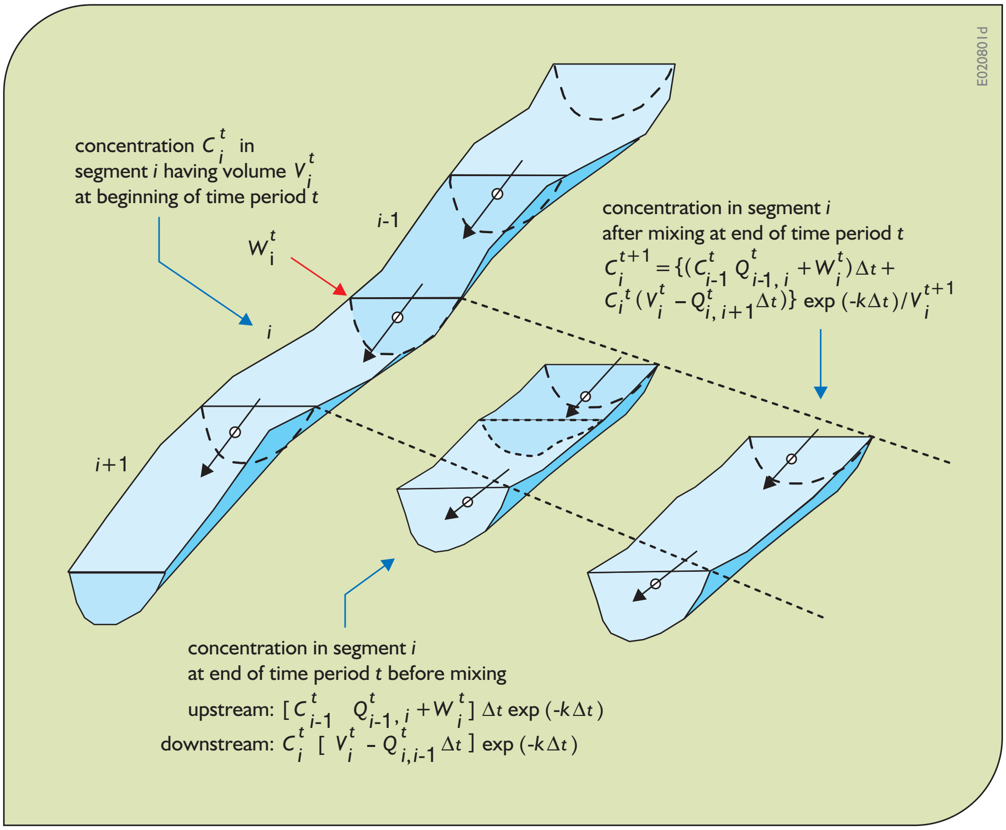



**Figure 12.26.** Water quality modelling approach showing a water system schematized into computational cells.

increasing flow velocities, also reduces the allowable duration of each time period $t$. The duration of each simulation time step $\Delta t$ must be such that flow from any segment or element enters only the adjacent downstream segment or element during that time step. Stated formally, the restriction is:

$$\Delta t \leq T_i \tag{12.117}$$

where $T_i$ is the residence time in reach segment or volume element $i$. For a 1-dimensional stream or river system consisting of a series of segments $i$ of length $\Delta x_i$, cross section area $A_i$ and average flow $Q_{it}$, the restriction is:

$$\Delta t \leq \min\{\Delta x_i A_i / Q_{it};\quad \forall i, t\} \tag{12.118}$$

If time steps are chosen that violate this condition, then numerical solutions will be in error. The restriction defined by Equation 12.118 is often termed the 'courant condition'. It limits the maximum time-step value. Since the flows being simulated are not always known, this leads to the selection of very small time steps, especially in water bodies having very little dispersion. While smaller simulation time steps increase the accuracy of the model output, they also increase the computational times. Thus the balance between computational speed and numerical accuracy restricts the model efficiency in the traditional approach to simulating water quality.

## 6.3. Backtracking Approach

An alternative Lagrangian or backtracking approach to water quality simulation eliminates the need to consider the simulation time-step duration restriction indicated by Equation 12.118 (Manson and Wallis, 2000; Yin, 2002). The backtracking approach permits any simulation time-step duration to be used along with any segmenting scheme. Unlike the traditional approach, water can travel through any number of successive segments or volume elements in each simulation time step.

This approach differs from the traditional one in that, instead of following the water in a segment or volume element downstream, the system tracks back upstream to find the source concentrations of the contaminants at time $t$ that will be in the control volume or segment $i + 1$ at the beginning of time period $t + 1$.

The backtracking process works from upstream to downstream. It starts from the segment of interest, $i$, and finds all the upstream sources of contaminants that flow into segment $i$ during time period $t$. The contaminants could come from segments in the same river reach or storage site, or from upstream river reaches or storage volume segments. They could also come from incremental flows into upstream segments. Flows between the source site and the segment $i+1$ transport the contaminants from their source sites to segment $i$ during the time interval $\Delta t$, as shown in Figure 12.27.

The simulation process for each segment and for each time period involves three steps. To compute the concentration of each constituent in segment $i$ at the end of time period $t$, as shown in Figure 12.27, the approach first backtracks upstream to locate all the contaminant particles at the beginning of period $t$ that will be in the segment $i$ at the end of period $t$. This is achieved by finding the most upstream and downstream positions of all reach intervals that will be at the corresponding boundaries of segment $i$ at the end of time period $t$. To do this requires computing the velocities through each of the intermediate segments or volume elements. Second, the changes in the amounts of the modelled quality constituents, such as temperature, organics, nutrients and toxics, are calculated, assuming plug flow during the time interval, $\Delta t$, and using the appropriate differential equations and numerical methods for solving them. Finally, all the multiple incoming blocks of water with their end-of-period constituent concentrations are completely mixed in the segment $i$ to obtain initial concentrations in that segment for the next time step, $t+1$. This is done for each segment $i$ in each time period $t$, proceeding in the downstream direction.

If no dispersion is assumed, the backtracking process can be simplified to consider only the end points of each reach. Backtracking can take place to each end-of-reach location whose time of travel to the point of interest is just equal to or greater than $\Delta t$. Then, using interpolation between end-of-period constituent concentrations at those upstream sites, plus all loadings between those sites and the downstream site of interest, the constituent

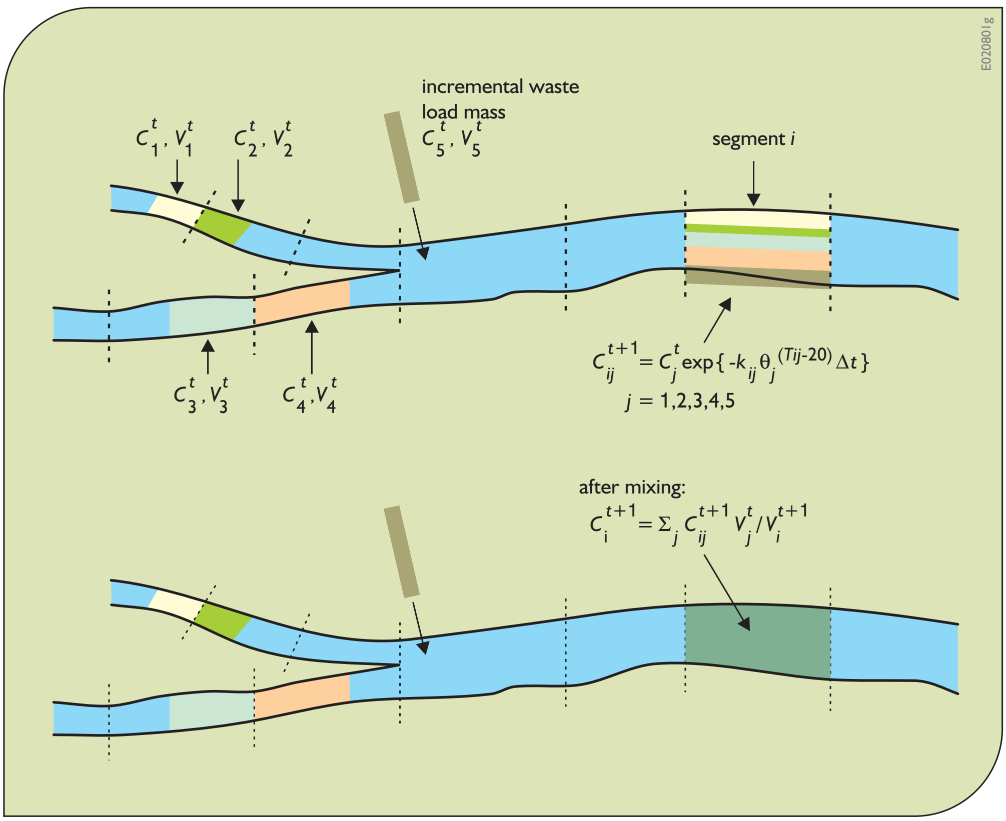



**Figure 12.27.** The backtracking approach for computing the concentrations of constituents in each reach segment or volume element $i$ during time-step duration of $\Delta t$.

concentrations at the end of the time period $t$ at the downstream ends of each reach can be computed. This process, like the one involving fully mixed reach segments, must take into account the possibility of multiple paths from each pollutant source to the site of interest, and the different values of rate constants, temperatures and other water quality parameters in each reach along those paths.

Figure 12.27 illustrates an example of backtracking involving simple first-order decay processes. Assume contaminants that end up in reach segment $i$ at the beginning of period $t + 1$ come from $J$ sources with initial concentrations $C_1^t, C_2^t, C_3^t, \ldots, C_j^t$ at the beginning of time period $t$. Decay of mass from each source $j$ during time $\Delta t$ in each segment or volume element is determined by the following differential equation:

$$dC_j^t/dt = -k_j \theta_j^{(T-20)} C_j^t \tag{12.119}$$

The decay rate constant $k_j$, temperature correction coefficient $\theta_j$ and water temperature $T$ are all temporally and spatially varied variables. Their values depend on the particular river reaches and storage volume sites through which the water travels during the period $t$ from sites $j$ to segment $i$.

Integrating Equation 12.119 yields:

$$C_j^{t+1} = C_j^t \exp\{-k_j \theta_j^{(T-20)} \Delta t\} \tag{12.120}$$

Since $\Delta t$ is the time it takes water having an initial concentration $C_j^t$ to travel to reach $i$, the values $C_j^{t+1}$ can be denoted as $C_{ij}^{t+1}$:

$$C_{ij}^{t+1} = C_j^t \exp\{-k_{ij} \theta_j^{(T_{ij}-20)} \Delta t\} \tag{12.121}$$

In Equation 12.121 the values of the parameters are the appropriate ones for the stream or river between the source segments $j$ and the destination segment $i$. These concentrations times their respective volumes, $V_j^t$, can then be mixed together to define the initial concentration, $C_i^{t+1}$, in segment $i$ at the beginning of the next time period $t + 1$.

## 6.4. Model Uncertainty

There are two significant sources of uncertainty in water quality management models. One stems from incomplete knowledge or lack of sufficient data to estimate the probabilities of various events that might happen. Sometimes it is difficult to even identify possible future events. This type of uncertainty – sometimes called epistemic (Stewart, 2000) – stems from our incomplete conceptual understanding of the systems under study, by models that are necessarily simplified representations of the complexity of the natural and socio-economic systems, as well as by limited data for testing hypotheses and/or simulating the systems.

Limited conceptual understanding leads to parameter uncertainty. For example, there is an ongoing debate about the parameters that can best represent the fate and transfer of pollutants through watersheds and water bodies. Arguably, more complete data and more work on model development can reduce this uncertainty. Thus, a goal of water quality management should be to increase the availability of data, improve their reliabilities and advance our modelling capabilities.

However, even if it were possible to eliminate knowledge uncertainty, complete certainty in support of water quality management decisions will probably never be achieved until we can predict the variability of natural processes. This type of uncertainty arises in systems characterized by randomness. Assuming past observations are indicative of what might happen in the future and with the same frequency – in other words, assuming stationary stochastic processes – we can estimate from these past observations the possible future events or outcomes that could occur and their probabilities. Even if we think we can estimate how likely any possible type of event may be in the future, we cannot predict precisely when or to what extent that event will occur.

For ecosystems, we cannot be certain we know even what events may occur in the future, let alone their probabilities. Ecosystems are open systems in which it is not possible to know in advance what all the possible biological outcomes will be. Surprises are not only possible, but likely; hence, neither type of uncertainty – knowledge uncertainty nor unpredictable variability or randomness – can be eliminated.

Thus, uncertainty is a reality of water quantity and quality management. This must be recognized when considering the results of water quality management models that relate to actions taken to meet the desired water quality criteria and designated uses of water bodies. Chapter 9 suggests some ways of characterizing this uncertainty.

# 7. Conclusions: Implementing a Water Quality Management Policy

This chapter has provided only a brief introduction to some of the relationships contained in water quality models. As can be said for the other chapters as well, it summarizes a subject on which entire texts, and very good ones, have been written (see, for example, Chapra, 1997; McCutcheon, 1989; Orlob, 1983; Schnoor, 1996; Thomann and Mueller, 1987). Water quality modelling and management demand skill and data. Skill comes with experience. Sufficient expertise will not be gained by working just with the material introduced in this chapter. It serves only as an introduction to surface water quality models, their assumptions and their limitations.

If accompanied by field data and uncertainty analysis, many existing models can be used to assist those responsible for developing water quality management plans in an adaptive implementation or management framework. Adaptive implementation or management will allow for both model and data improvements over time. Such approaches strive toward achieving water quality standards while relying on monitoring and experimentation to reduce uncertainty. This is often the only way one can proceed, given the complexity of the real world compared to the predictive models and the data and time usually available at the time a water quality analysis is needed. Starting with simple analyses and iteratively expanding data collection and modelling as the need arises is a reasonable approach.

An adaptive management process begins with initial actions that have reasonable chances of succeeding. Future actions must be based on continued monitoring of the water body to determine how it responds to the actions taken. Plans for future regulation and public spending should be subject to revision as stakeholders learn more about how the system responds to actions taken. Monitoring is an essential aspect of adaptive water quality management and modelling (see Appendix B).

Regardless of what immediate actions are taken, there may not be an immediate measurable response. For example, there may be significant lags between the time when actions are taken to reduce nutrient loads and the resulting changes in nutrient concentrations. This is especially likely if nutrients from past activities are tightly bound to sediments or if nutrient-contaminated groundwater has a long residence time before its release to surface water. For many reasons, lags between actions taken and responses must be expected. Water bodies should be monitored to establish whether the 'trajectories' of the measured water quality criteria point toward attainment of the designated use.

Wasteload allocations will inevitably be required if quality standards are not being met. These involve costs. Different allocations will have different total costs and different distributions of those costs; hence they will have different perceived levels of fairness. A minimum-cost policy may result in a cost distribution that places most of the burden on just some of the stakeholders. But until such a policy is identified, one will not know this. An alternative may be to reduce loads from all sources by the same proportion. Such a policy has prevailed in the United States over the past several decades. Even though not very cost-effective from the point of view of water quality management, the ease of administration and the fulfilment of other objectives must have made such a policy politically acceptable, even though expensive. However, these types of wasteload allocations policies will not in themselves be sufficient for many of the ecosystem restoration efforts that are increasingly being made. Restoration activities are motivated in part by the services ecosystems provide for water quality management.

Our abilities to include ecosystem components within water quantity and quality management models are at a fairly elementary level. Given the uncertainty, especially with respect to the prediction of how ecosystems will respond to water management actions, together with the need to take actions now, long before we can improve these capabilities, the popular call is for adaptive management. The trial and error aspects of adaptive management based on monitoring and imperfect models may not satisfy those who seek more definitive direction from water quality analysts and their predictive models. Stakeholders and responsible agencies seeking assurances that the actions taken will always work as predicted may be disappointed. Even the best predictive capabilities of science cannot ensure that an action that will lead to the attainment of designated uses will be initially identified. Adaptive management is the only reasonable option in most cases for allowing water quality management programmes to move forward in the face of considerable uncertainties.

## 8. References

AMBROSE, R.B.; BARNWELL, T.O.; MCCUTCHEON, S.C. and WILLIAMS, J.R. 1996. Computer models for water quality analysis. In: L.W. Mays (ed.), *Water Resources Handbook*. New York, McGraw-Hill. Chapter 14.

BORSUK, M.E.; STOW, C.A. and RECKHOW, K.H. 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*, Vol. 173, pp. 219–39.

BROWN, L.C. and BARNWELL, T.O. Jr. 1987. *The enhanced stream water-quality models QUAL2E and QUAL2E-UNCAS: documentation and user manual*. EPA-600/3-87/007. Athens, Ga., EPA Environmental Research Laboratory.

CERCO, C.F. and COLE, T. 1995. *User's guide to the CE-QUAL-ICM three-dimensional eutrophication model, release version 1.0*. Technical Report EL-95-15. Vicksburg, Miss., USACE Waterways Experiment Station.

CHAPRA, S. 1997. *Surface water quality modelling*. New York, McGraw-Hill.

DEMARCHI, C.; IVANOV, P.; JOLMA, A.; MASLIEV, I.; SMITH, M. and SOMLYÓDY, L. 1999. Innovative tools for water quality management and policy analysis: DESERT and STREAMPLAN. *Water Science and Technology*, Vol. 40, No. 10, pp. 103–10.

ELMORE, H.L. and HAYES, T.W. 1960. Solubility of atmospheric oxygen in water. *Journal of the Environmental Engineering Division*, ASCE, Vol. 86, No. SA4, pp. 41–53.

ENGELUND, F. and HANSEN, E. 1967. A monograph on sediment transport in alluvial streams. Copenhagen, Teknisk forlag.

IVANOV, P.; MASLIEV, I.; DE MARCHI, C. and SOMLYÓDY, L. 1996. DESERT: decision support system for evaluating river basin strategies; user's manual. *International Institute for Applied Systems Analysis*. Laxenburg, Austria, IISA.

KARR, J.R. 1990. Bioassessment and non-point source pollution: an overview. In: *Second National Symposium on Water Quality Assessment*. Washington, D.C., EPA Office of Water; pp. 4–18.

KRONE, R.B. 1962. *Flume studies of the transport of sediment in estuarial shoaling processes*. Berkeley, Calif., University of California, Hydraulic and sanitary engineering laboratory.

LOS, F.J. 1991. *Mathematical simulation of algae blooms by the model BLOOM II, Version 2, T68*. Delft, the Netherlands, WL | Delft Hydraulics Report.

LOS, F.J. et al. 1992. *Process formulations DBS. WL | Delft Hydraulics, model documentation T542*. Delft, the Netherlands, Delft Hydraulics, (In Dutch.)

MANSON, J.R. and WALLIS, S.G. 2000. A conservative semi-lagrangian fate and transport model for fluvial systems – I. Theoretical Development. *Journal of Water Research*, Vol. 34, No. 15, pp. 3769–77.

MCCUTCHEON, S.C. 1989. *Water Quality Modelling, Vol. 1*. Boca Raton, Fla., CRC.

MOLEN, D.T. VAN DER; LOS, F.J.; VAN BALLEGOOIJEN, L. and VAN DER VAT, M.P. 1994. Mathematical modelling as a tool for management in eutrophication control of shallow lakes. *Hydrobiologia*, Vol. 275–6, pp. 479–92.

MONBALIU, J.; HARGREAVES, J.C.; CARRETERO, J.C.; GERRITSEN, H. and FLATHER, R.A. 1999. Wave modelling in the PROMISE project. *Coastal Engineering,* Vol. 37, Nos. 3–4, pp. 379–407.

NRC (National Research Council). 1992. *Restoration of aquatic ecosystems*. Washington, DC, National Academy Press.

NRC (National Research Council). 2001. *Assessing the TMDL approach to water quality management committee to assess the scientific basis of the total maximum daily load approach to water pollution reduction*. Washington, D.C., Water Science and Technology Board, National Academy Press.

ORLOB, G.T. (ed.). 1983. Mathematical modelling of water quality: streams, lakes and reservoirs. Chichester, UK, Wiley.

PARTHENIADES, E. 1962. A study of erosion and deposition of cohesive soils in salt water. Berkeley, Calif., University of California.

REICHERT, P. 1994. AQUASIM: a tool for simulation and data analysis of aquatic systems. *Water Science and Technology,* Vol. 30, No. 2, pp. 21–30.

SCHNOOR, J.L. 1996. *Environmental modelling, fate and transport of pollutants in water, air and soil*. New York, Wiley.

SMITS, J. 2001. *DELWAQ-BLOOM-switch*. Delft, Delft Hydraulics.

STEWART, T.R. 2000. Uncertainty, judgment, and error in prediction. In: D. Sarewitz, R.A. Pielke Jr. and R. Byerly Jr. (eds.), *Prediction: science, decision-making, and the future of nature*. Washington, D.C., Island Press.

STREETER, H.W. and PHELPS, E.B. 1925. *A study of the pollution and natural purification of the Ohio River, III. Factors concerned in the phenomena of oxidation and reaeration*. Washington, D.C., US Public Health Service.

THOMANN, R.V. and MUELLER, J.A. 1987. *Principles of surface water quality modelling and control*. New York, Harper and Row.

USEPA. 2001. *BASINS Version 3.0 User's Manual*. EPA-823-B-01-001. Washington, D.C., EPA Office of Water and Office of Science and Technology.

WL | DELFT HYDRAULICS. 2003. *Delft3D-WAQ, user manual; versatile water quality modelling in 1D, 2D, or 3D systems including physical, (bio)chemical and biological processes*. Delft, the Netherlands, WL | Delft Hydraulics.

YIN, H. 2002. *Development of a watershed information system*, M.Sc. Thesis. Ithaca, N.Y., Cornell University, Civil and Environmental Engineering Department.

## Additional References (Further Reading)

ASCE. 1999. *National stormwater best management practices (BMP) data-base: version 1.0*. User's Guide and CD. Washington, D.C. Prepared by Urban Water Resources Research Council of ASCE, and Wright Water Engineers, Inc., Urban Drainage and Flood Control District, and URS Greiner Woodward Clyde, in cooperation with EPA Office of Water.

BECK, M.B. 1987. Water quality modelling: a review of the analysis of uncertainty. *Water Resources Research,* No. 23, pp. 1393–442.

BECK, M.B. and VAN STRATEN, G. (eds.). 1983. *Uncertainty and forecasting of water quality*. Berlin, Springer-Verlag.

BISWAS, A.K. (ed.). 1997. *Water resources: environmental planning, management, and development*. New York, McGraw-Hill.

BOWIE, G.L.; MILLS, W.B.; PORCELLA, D.B.; CAMPBELL, C.L.; PAGENKOPF, J.R.; RUPP, G.L.; JOHNSON, K.M.; CHAN, P.W.H.; GHERINI, S.A. and CHAMBERLIN, C.E. 1985. *Rates, constants, and kinetics: formulations in surface water quality modelling*, 2nd edn. Report EPA/600/3-85/040. Athens, Ga., US EPA.

CHURCHILL, M.A.; ELMORE, H.L. and BUCKINGHAM, R.A. 1962. Prediction of stream reaeration rates. *Journal of the Sanitary Engineering Division*, ASCE Vol. 88, No. SA4, No. 1, pp. 1–46.

DELVIGNE, G.A.L. 1980. *Natural reaeration of surface water*. WL | Delft Hydraulics, Report on literature study R1149. Delft, WL | Delft Hydraulics, (In Dutch.)

DITORO, D.M.; PAQUIN, P.R.; SUBBURAMU, K. and GRUBER, D.A. 1990. Sediment oxygen demand model: methane and ammonia oxidation. *Journal of Environmental Engineering*, ASCE, Vol. 116, No. 5, pp. 945–86.

DOBBINS, W.E. 1964. BOD and oxygen relationships in streams. *Journal of the Sanitary Engineering Division*, ASCE, Vol. 90, No. SA3, pp. 53–78.

European Union. 2001a. *The EU water framework directive*. http://europa.eu.int/water/water-framework/index_en.html. Accessed 21 June.

European Union. 2001b. *Water protection and management: framework directive in the field of water policy*. http://www.europa.eu.int/scadplus/leg/en/lvb/l28002b.htm. Accessed 20 June.

European Union. 2001c. *Water protection and management: urban waste water treatment*. http://www.europa.eu.int/scadplus/leg/en/lvb/l28008.htm. Accessed 20 June.

FAGERBAKKE, K.M.; HELDAL, M. and NORLAND, S. 1996. Content of carbon, nitrogen, oxygen, sulfur and phosphorus in native aquatic and cultured bacteria. *Aquatic Microbial Ecology,* No. 10, pp. 15–27.

GROMIEC, M.J.; LOUCKS, D.P. and ORLOB, G.T. 1982. Stream quality modelling. In: G.T. Orlob (ed.), *Mathematical modelling of water quality*. Chichester, UK, Wiley, pp. 176–226.

HORNBERGER, G.M. and SPEAR, R.C. 1981. An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management,* No. 12, pp. 7–18.

HORNBERGER, G.M. and SPEAR, R.C. 1983. An approach to the analysis of behaviour and sensitivity in environmental systems. In: M.B. Beck and G. van Straten (eds.), *Uncertainty and forecasting of water quality*, Berlin, Springer-Verlag, pp. 101–16.

KARR, J.R. 2000, Health, integrity, and biological assessment: the importance of whole things. In: D. Pimentel, L. Westra, and R.F. Noss (eds.), *Ecological integrity: integrating environment, conservation, and health*, Washington, D.C., Island Press, pp. 209–26. and KARR, J.R.; CHU, E.W. Sustaining living rivers. *Hydrobiologia*, Nos. 422–3, pp. 1–14.

KARR, J.R. and DUDLEY, D.R. 1981. Ecological perspective on water quality goals. *Environmental Management,* No. 5, pp. 55–68.

LANGBIEN, W.B. and DURUM, W.H. 1967. The aeration capacity of streams, Circular 542. Washington, D.C., USGS.

MAIDMENT, D.R. (ed.). 1993. *Handbook of hydrology*. New York, McGraw-Hill.

MASLIEV, I.; SOMLYÓDY, L. and KONCSOS, L. 1995. *On reconciliation of traditional water quality models and activated sludge models*. Working paper WP 95–18. Laxenburg, Austria, International Institute for Applied Systems Analysis.

MILLS, W.B.; PORCELLA, D.B.; UNGS, M.J.; GHERINI, S.A.; SUMMERS, K.V.; MOK, L.; RUPP, G.L.; BOWIE, G.L. and HAITH, D.A. 1985. *Water quality assessment: a screening procedure for toxic and conventional pollutants in surface and ground water, Parts I and II*. EPA/600/6-85/002a,b. Athens, Ga., Environmental Protection Agency.

MORGAN, M.G. and HENRION, M. 1990. *Uncertainty*. New York: Cambridge University Press.

NATIONAL ACADEMY OF PUBLIC ADMINISTRATION. 2000. *Transforming environmental protection for the twenty-first century*. Washington, D.C., National Academy of Public Administration.

NOVOTNY, V. 1999. Integrating diffuse/nonpoint pollution control and water body restoration into watershed management. *Journal of the American Water Resources Association*, Vol. 35, No. 4, pp. 717–27.

NOVOTNY, V. and OLEM, H. 1994. Water quality: prevention, identification and management of diffuse pollution. New York, Van Nostrand-Reinhold (distributed by Wiley). Ohio, Ohio EPA.

O'CONNOR, D.J. 1961. Oxygen balance of an estuary. *J. San. Eng. Div. ASCE,* Vol. 86, No. SA3, pp. 35–55.

O'CONNOR, D.J. and DOBBINS, W.E. 1958. Mechanism of reaeration in natural streams. *Transactions of the American Society of Civil Engineers,* No. 123, pp. 641–66.

OWENS, M.; EDWARDS, R.W. and GIBBS, J.W. 1964. Some reaeration studies in streams. *International Journal of Air and Water Pollution,* No. 8, pp. 469–86.

PETERS, R.H. 1991. A critique for ecology. Cambridge, Cambridge University Press.

REICHERT, P. 1995. Design techniques of a computer program for the identification of processes and the simulation of water quality in aquatic systems. *Environmental Software and Modeling,* Vol. 10, No. 3, pp. 199–210.

REICHERT, P. 2001. River Water-quality model no. 1 (RWQM1): case study II. Oxygen and nitrogen conversion processes in the River Glatt (Switzerland). *Water Science and Technology,* Vol. 43, No. 5, pp. 51–60.

REICHERT, P.; BORCHARDT, D.; HENZE, M.; RAUCH, W.; SHANAHAN, P.; SOMLYÓDY, L. and VANROLLEGHEM, P. 2001. River water-quality model No. 1 (RWQM1): II. Biochemical process equations. *Wat. Sci. Tech.,* Vol. 43, No. 5, pp. 11–30.

REICHERT, P. and VANROLLEGHEM, P. 2001. Identifiability and uncertainty analysis of the river water-quality model No. 1 (RWQM1). *Water Science and Technology,* Vol. 43, No. 7, pp. 329–38.

ROESNER, L.A.; GIGUERE, P.R. and EVENSON, D.E. 1981. *Computer program documentation for the stream quality model QUAL–II*. Report EPA 600/9–81–014. Athens, Ga., USEPA.

SHANAHAN, P.; BORCHARDT, D.; HENZE, M.; RAUCH, W.; REICHERT, P.; SOMLYÓDY, L. and VANROLLEGHEM, P. 2001. River Water-quality model no. 1 (RWQM1): I Modelling approach. *Water Science and Technology,* Vol. 43, No. 5, pp. 1–9.

SHANAHAN, P.; HENZE, M.; KONCSOS, L.; RAUCH, W.; REICHERT, P.; SOMLYÓDY, L. and VANROLLEGHEM, P. 1998. River water quality modelling: II. Problems of the art. *Water Science and Technology,* Vol. 38, No. 11, pp. 245–52.

SHEN, H.W. and JULIEN, P.Y. 1993. Erosion and sediment transport. In: D.R. Maidment (ed.), *Handbook of hydrology*, Chapter 12. New York, McGraw-Hill.

SMITH, R.A.; SCHWARZ, G.E. and ALEXANDER, R.B. 1997. Regional interpretation of water quality monitoring data. *Water Resources Research,* Vol. 33, No. 12, pp. 2781–98.

SOMLYÓDY, L. 1982. Water quality modelling: a comparison of transport-oriented and ecology-oriented approaches. *Ecological Modelling,* No. 17, pp. 183–207.

SOMLYÓDY, L.; HENZE, M.; KONCSOS, L.; RAUCH, W.; REICHERT, P.; SHANAHAN, P. and VANROLLEGHEM, P. 1998. River water quality modelling – III. Future of the art. *Water Science and Technology,* Vol. 38, No. 11, pp. 253–60.

SOMLYÓDY, L. and VAN STRATEN, G. (eds.). 1986. *Modelling and managing shallow lake eutrophication, with application to Lake Balaton*. Berlin, Springer-Verlag.

SOMLYÓDY, L. and VARIS, O. 1993. Modelling the quality of rivers and lakes. In: P. Zannetti (ed.), *Environmental modelling: computer methods and software for simulating environmental pollution and its adverse effects*, Vol. 1, pp. 213–57. Southampton, UK, CMP.

SPANOU, M. and CHEN, D. 2001. Water quality modelling of the Upper Mersey River system using an object-oriented framework. *Journal of Hydroinformatics*, Vol. 3, No. 3, pp. 173–94.

SPEAR, R. and HORNBERGER, G.M. 1980. Eutrophication in Peel Inlet – II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research,* Vol. 14, pp. 43–9.

STUMM, W. and MORGAN, J.J. 1981. *Aquatic Chemistry*. New York, Wiley.

THOMAS, H.A. Jr. 1948. Pollution load capacity of streams. *Water and Sewage Works*, No. 95, p. 409.

TOXOPEUS, A.G. 1996. *An interactive spatial and temporal modelling system as a tool in ecosystem management*. Enschede, the Netherlands, International Institute for Aerospace Survey and Earth Sciences, ITC.

ULANOWICZ, R.E. 1997. *Ecology, the ascendant perspective*. New York, Columbia University Press.

USEPA. 1993. *The watershed protection approach: the annual report 1992*. EPA 840-S-93-001. Washington, D.C., EPA Office of Water.

USEPA. 1994. *Water quality standards handbook*, 2nd edn. EPA 823-B-94-005a. Washington, D.C., EPA Office of Water.

USEPA. 1995a. *Environmental indicators of water quality in the United States*. EPA 841-R-96-002. Washington, D.C., Office of Policy, Planning, and Evaluation.

USEPA. 1995b. *A conceptual framework to support development and use of environmental information in decision-making*. EPA 239-R-95-012. Washington, D.C., Office of Policy, Planning, and Evaluation.

USEPA. 1995c. *Ecological restoration: a tool to manage stream quality*. Report EPA 841-F-95-007. Washington, D.C., USEPA.

USEPA. 1998. *Lake and Reservoir bioassessment and biocriteria: technical guidance document*. EPA 841-B-98-007. Washington, DC: EPA Office of Water.

USEPA. 1999. *Draft guidance for water quality-based decisions: the TMDL process*, 2nd edn. Washington, D.C., EPA Office of Water.

USEPA. 2000. *Stressor identification guidance document*. EPA-822-B-00-025. Washington, D.C., EPA Office of Water and Office of Research and Development.

VAN PAGEE, J.A. 1978. *Natural reaeration of surface water by the wind*. Report on literature study R1318-II. Delft, the Netherlands, WL | Delft Hydraulics. (In Dutch.)

VAN RIJN, L.C. 1984. Bed load transport (part I), suspended load transport (part II). *Journal of Hydraulic Engineering*, Vol. 110, Nos. 10–1, pp. 1431–56, 1613–41.

VAN STRATEN, G. 1983. Maximum likelihood estimation of parameters and uncertainty in phytoplankton models. In: M.B. Beck and G. van Straten (eds.), *Uncertainty and forecasting of water quality*. Berlin, Springer-Verlag.

YOUNG, P. 1998. Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling and Software*, Vol. 13, pp. 105–22.

YOUNG, W.J.; LAM, D.C.L.; RESSEL, V. and WONG, J.W. 2000. Development of an Environmental Flows Decision support system. *Environmental Modelling and Software*, Vol. 15, pp. 257–65.

# 13. Urban Water Systems


1. Introduction 427

2. Drinking Water 428
   2.1. Water Demand 428
   2.2. Water Treatment 428
   2.3. Water Distribution 430
      2.3.1. Open Channel Networks 432
      2.3.2. Pressure Pipe Networks 432
      2.3.3. Water Quality 434

3. Wastewater 434
   3.1. Wastewater Production 434
   3.2. Sewer Networks 434
   3.3. Wastewater Treatment 435

4. Urban Drainage 437
   4.1. Rainfall 437
      4.1.1. Time Series Versus Design Storms 437
      4.1.2. Spatial-Temporal Distributions 438
      4.1.3. Synthetic Rainfall 438
      4.1.4. Design Rainfall 438
   4.2. Runoff 439
      4.2.1. Runoff Modelling 439
      4.2.2. The Horton Infiltration Model 441
      4.2.3. The US Soil Conservation Method (SCS) Model 442
      4.2.4. Other Rainfall–Runoff Models 444
   4.3. Surface Pollutant Loading and Washoff 445
      4.3.1. Surface Loading 446
      4.3.2. Surface Washoff 446
      4.3.3. Stormwater Sewer and Pipe Flow 447
      4.3.4. Sediment Transport 448
      4.3.5. Structures and Special Flow Characteristics 448
   4.4. Water Quality Impacts 448
      4.4.1. Slime 448
      4.4.2. Sediment 448
      4.4.3. Pollution Impact on the Environment 448
      4.4.4. Bacteriological and Pathogenic Factors 451
      4.4.5. Oil and Toxic Contaminants 451
      4.4.6. Suspended Solids 452

5. Urban Water System Modelling 452
   5.1. Model Selection 452
   5.2. Optimization 453
   5.3. Simulation 455

6. Conclusions 456

7. References 457

# 13 Urban Water Systems

'Today, a simple turn of the tap provides clean water – a precious resource. Engineering advances in managing this resource – with water treatment, supply, and distribution systems – changed life profoundly in the twentieth century, virtually eliminating waterborne diseases in developed nations, and providing clean and abundant water for communities, farms, and industries'.

So states the US National Academy of Engineering on its selection of water supply systems to be among the five greatest achievements of engineering in the twentieth century. As populations continually move to urban areas for improved opportunities and a higher standard of living, and as cities merge to form megacities, the design and management of water supply systems serving these urban areas becomes an increasingly important part of regional integrated water resources planning and management.

## 1. Introduction

Urban water infrastructure typically includes water collection and storage facilities at source sites, water transport via aqueducts (canals, tunnels and/or pipelines) from source sites to water treatment facilities; water treatment, storage and distribution systems; wastewater collection (sewage) systems and treatment; and urban drainage works. This is illustrated as a simple schematic in Figure 13.1.

Generic simulation models of components of urban water systems have been developed and are commonly applied to study specific component design and operation issues. Increasingly, optimization models are being used to estimate cost-effective designs and operating policies. Cost savings can be substantial, especially when applied to large complex urban systems (Dandy and Engelhardt, 2001; Savic and Walters, 1997).

Most urban water users require high-quality water, and natural surface and/or groundwater supplies, called raw water, often cannot meet the quality requirements of domestic and industrial users. In such situations, water treatment is required prior to its use. Once it is treated, urban water can then be stored and distributed within the urban area, usually through a network of storage tanks and pipes. Pipe flows in urban distribution systems should be under pressure to prevent contamination from groundwaters and to meet various user and fire-protection requirements.

After use, the 'wastewater' is collected in a network of sewers, or in some cases ditches, leading to a wastewater treatment plant or discharge site. In many urban areas the sewage system has a dual function. The sewers collect both wastewater from households and the runoff from streets and roofs during storm events. However, the transport capacity of the sewer network and the treatment facilities are limited. During intense rainfall, overflows from the sewage system discharge a mixture of surface runoff and wastewater to the surface waters. This has a negative impact on the water quality of urban surface waters.

Wastewater treatment plants remove some of the impurities in the wastewater before it is discharged into receiving water bodies or on land surfaces. Water bodies receiving effluents from point sources such as wastewater treatment plants may also receive runoff from the surrounding watershed area during storm events. The pollutants in both point and non-point discharges will

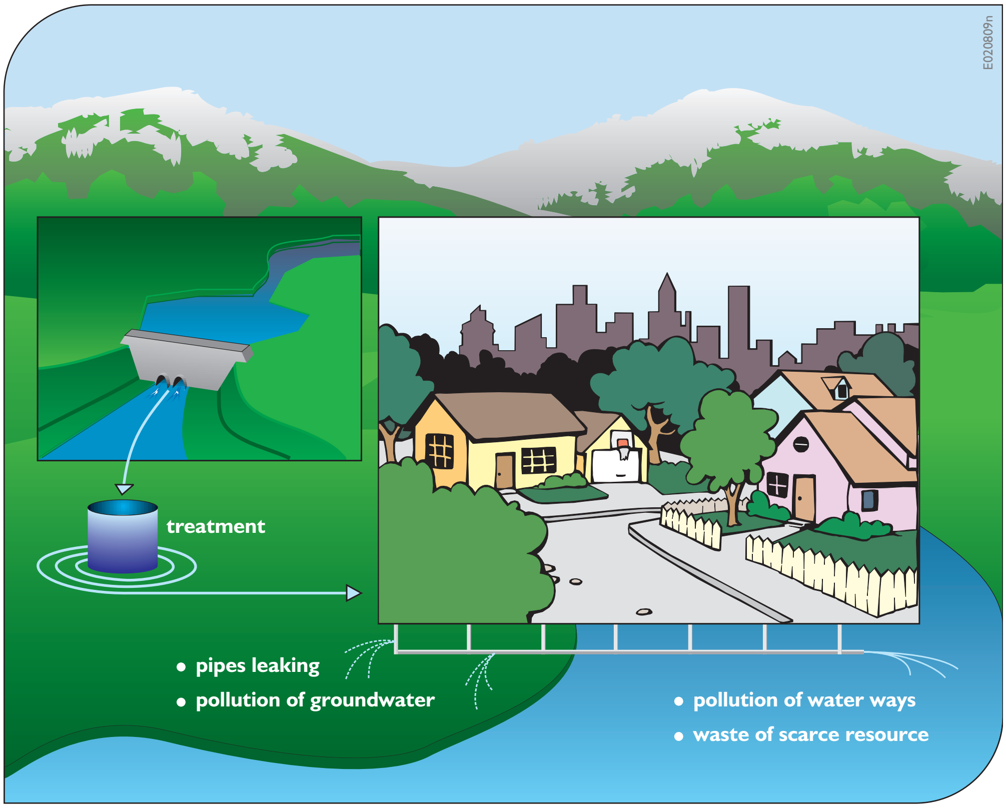



**Figure 13.1.** Schematic showing urban surface water source, water treatment prior to urban use, and some sources of non-point urban drainage and runoff and its impacts.

affect the quality of the water in those receiving water bodies. The fate and transport of these pollutants in these water bodies can be predicted by using water quality models similar to those discussed in Chapter 12.

This chapter briefly describes these urban water system components and reviews some of the general assumptions incorporated into optimization and simulation models used to plan urban water systems. The focus of urban water systems modelling is mainly on the prediction and management of quantity and quality of flows and pressure heads in water distribution networks, wastewater flows in gravity sewer networks, and on the design efficiencies of water and wastewater treatment plants. Other models can be used for the real-time operation of various components of urban systems.

# 2. Drinking Water

Drinking water issues include demand estimation, water treatment, and distribution.

## 2.1. Water Demand

A secure water supply is of vital importance for the health of the population and for the economy. Drinking water demand depends on:

- the number of inhabitants with access to drinking water
- meteorological and climatological conditions
- the price of drinking water
- the availability of drinking water
- an environmental policy that aims at moderate use of drinking water.

Table 13.1 shows an overview of the total annual water demand in various countries. The total water demand is sub-divided into domestic use and agricultural and industrial water use.

Drinking water demand for domestic use shows a daily and seasonal variation. There is no general formula for predicting drinking water demand. Drinking water suppliers tend to make predictions on the basis of their own experience and historical information about water demand in their region.

| country | demand m$^3$/capita | domestic m$^3$/capita | agriculture m$^3$/capita | industrial m$^3$/capita | year |
|---|---|---|---|---|---|
| Germany | 490 | 67 | 2 | 389 | 1999 |
| USA | 1870 | 213 | 752 | 828 | 1990 |
| Mexico | 800 | 101 | 662 | 38 | 1999 |
| Egypt | 920 | 55 | 792 | 74 | 1993 |
| Namibia | 185 | 52 | 126 | 6 | 1990 |
| China | 439 | 22 | 338 | 78 | 1993 |
| India | 588 | 29 | 18 | 541 | 1990 |

source: OECD data compendium 2002 (Germany, USA, Mexico)

source: World Resources Institute (Egypt, Namibia, China, India)

E040712b

**Table 13.1.** Annual per-capita water demand in various countries in the world. Source: 1) OECD data compendium 2002 and 2) World Resources Institute.

## 2.2. Water Treatment

Before water is used for human consumption, its harmful impurities need to be removed. Communities that do not have adequate water treatment facilities, a common problem in developing regions, often have high incidences of disease and mortality due to drinking contaminated water. A range of syndromes, including acute dehydrating diarrhoea (cholera), prolonged febrile illness with abdominal symptoms (typhoid fever), acute bloody diarrhoea (dysentery) and chronic diarrhoea (Brainerd diarrhoea). Numerous health organizations point to the fact that contaminated water leads to over 3 billion episodes of diarrhoea and an estimated 2 million deaths, mostly among children, each year.

Contaminants in natural water supplies can also include microorganisms such as Cryptosporidium and Giardia lamblia as well as inorganic and organic cancer-causing chemicals (such as compounds containing arsenic, chromium, copper, lead and mercury) and radioactive material (such as radium and uranium). Herbicides and pesticides reduce the suitability of river water as a source of drinking water. Recently, traces of hormonal substances and medicines detected in river water are generating more and more concern.

To remove impurities and pathogens, a typical municipal water purification system involves a sequence of processes, from physical removal of impurities to chemical treatment. Physical and chemical removal processes include initial and final filtering, coagulation, flocculation, sedimentation and disinfection, as illustrated in the schematic of Figure 13.2.

As shown in Figure 13.2, one of the first steps in most water treatment plants involves passing raw water through coarse filters to remove sticks, leaves and other large solid objects. Sand and grit settle out of the water during this stage. Next a chemical such as alum is added to the raw water to facilitate coagulation. As the water is stirred, the alum causes the formation of sticky globs of small particles made up of bacteria, silt and other impurities. Once these globs of matter are formed, the water is routed to a series of settling tanks where the globs, or floc, sink to the bottom. This settling process is called flocculation.

After flocculation, the water is pumped slowly across another large settling basin. In this sedimentation or clarification process, much of the remaining floc and solid material accumulates at the bottom of the basin. The clarified water is then passed through layers of sand, coal and other granular material to remove microorganisms – including viruses, bacteria and protozoa such as Cryptosporidium – and any remaining floc and silt. This stage of purification mimics the natural filtration of water as it moves through the ground.

The filtered water is then treated with chemical disinfectants to kill any organisms that remain after the filtration process. An effective disinfectant is chlorine, but its use may cause potentially dangerous substances such as carcinogenic trihalomethanes.

Alternatives to chlorine include ozone oxidation (Figure 13.2). Unlike chlorine, ozone does not stay in the water after it leaves the treatment plant, so it offers

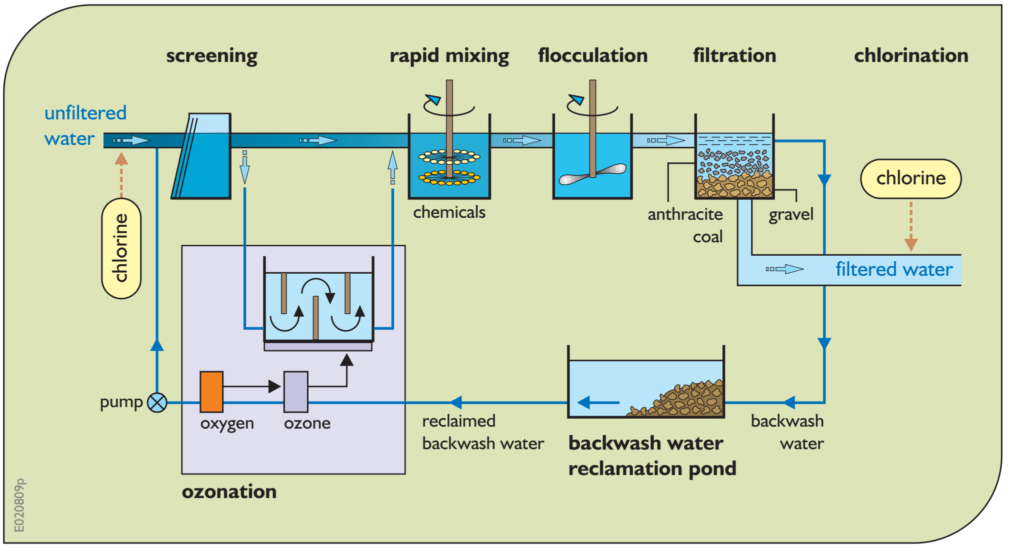



**Figure 13.2.** Typical processes in water treatment plants.

**Figure 13.3.** A 6-million gallon per day water treatment plant at San Luis Obispo, located about halfway between Los Angeles and San Francisco on the central coast of California.

no protection from bacteria that might be in the storage tanks and water pipes of the water distribution system. Water can also be treated with ultraviolet light to kill microorganisms, but this has the same limitation as oxidation: it is ineffective outside of the treatment plant.

Figure 13.3 is an aerial view of a water treatment plant serving a population of about 50,000.

Sometimes calcium carbonate is removed from drinking water in order to prevent it from accumulating in drinking water pipes and washing machines.

In arid coastal areas desalinated brackish or saline water is an important source of water for high-value uses. The cost of desalination is still high, but decreasing steadily. The two most common methods of desalination are distillation and reverse osmosis. Distillation requires more energy, while osmosis systems need frequent maintenance of the membranes.

## 2.3. Water Distribution

Water distribution systems include pumping stations, distribution storage and distribution piping. The hydraulic performance of each component depends upon the performance of the others. Of interest to designers are both the flows and their pressures. Leakage of drinking water from the distribution system is a concern in many old drinking water systems.

The energy at any point within a network of pipes is often represented in three parts: the pressure head, $p/\gamma$, the elevation head, $Z$, and the velocity head, $V^2/2g$. (A more precise representation includes a kinetic energy correction factor, but that factor is small and can be ignored.) For open-channel flows, the elevation head is the distance from some datum to the top of the water surface. For pressure-pipe flow, the elevation head is the distance from some datum to the centre of the pipe. The parameter $p$ is the pressure, for example Newtons per cubic metre (N/m$^3$), $\gamma$ is the specific weight (N/m$^2$) of water, $Z$ is the elevation above some base elevation (m), $V$ is the velocity (m/s), and $g$ is the gravitational acceleration (9.81 m/s$^2$).

Energy can be added to the system such as by a pump, or lost by, for example, friction. These changes in energy are referred to as head gains and losses. Balancing the energy across any two sites $i$ and $j$ in the system requires that the total heads, including any head gains $H_G$ and losses $H_L$ (m) are equal.

$$[p/\gamma + Z + V^2/2g + H_G]_{\text{site } i} = [p/\gamma + Z + V^2/2g + H_L]_{\text{site } j} \qquad (13.1)$$

The hydraulic grade is the sum of the pressure head and elevation head ($p/\gamma + Z$). For open-channel flow, the hydraulic grade is the water surface slope, since the pressure head at its surface is 0. For a pressure pipe, the hydraulic head is the height to which a water column would rise in a piezometer (a tube rising from the pipe). When plotted in profile along the length of the conveyance section, this is often referred to as the hydraulic grade line, or *HGL*. The hydraulic grade lines for open channels and pressure pipes are illustrated in Figures 13.4 and 13.5.

The energy grade is the sum of the hydraulic grade and the velocity head. This is the height to which a column of water would rise in a Pitot tube, but also accounts for fluid velocity. When plotted in profile, as in Figure 13.5, this is often referred to as the energy grade line, or *EGL*. At a lake or reservoir, where the velocity is essentially zero, the *EGL* is equal to the *HGL*.

Specific energy, $E$, is the sum of the depth of flow and the velocity head, $V^2/2g$. For open-channel flow, the depth of flow, $y$, is the elevation head minus the channel bottom elevation. For a given discharge, the specific energy is solely a function of channel depth. There may be more than one depth with the same specific energy. In one case the flow is subcritical (relatively higher depths, lower velocities) and in the other case the flow is critical (relatively lower depths and higher velocities). Whether or not the flow is above or below the critical depth (the depth that minimizes the specific energy) will depend in part on the channel slope.

Friction is the main cause of head loss. There are many equations that approximate friction loss associated with fluid flow through a given section of channel or pipe. These include Manning's or Strickler's equation, which is commonly used for open-channel flow, and Chezy's or Kutter's equation, Hazen–Williams equation, and

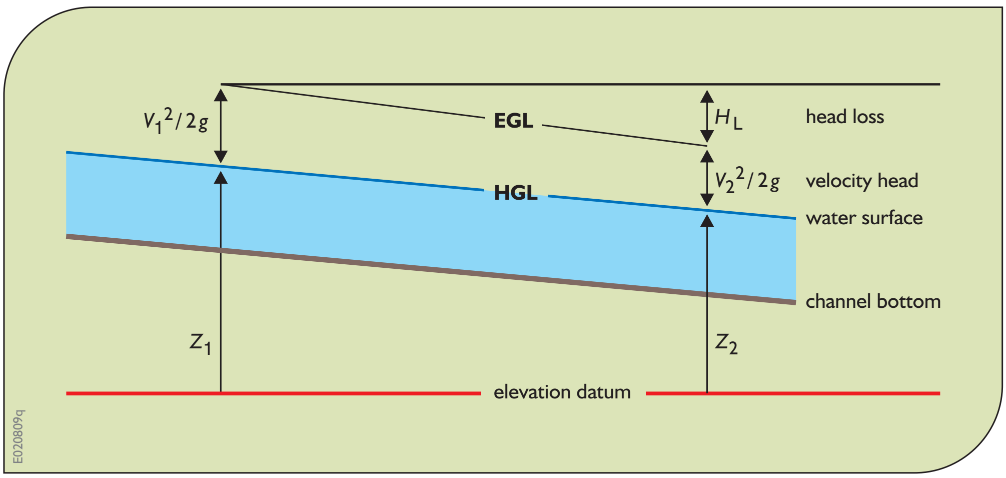



**Figure 13.4.** The energy components along an open channel.

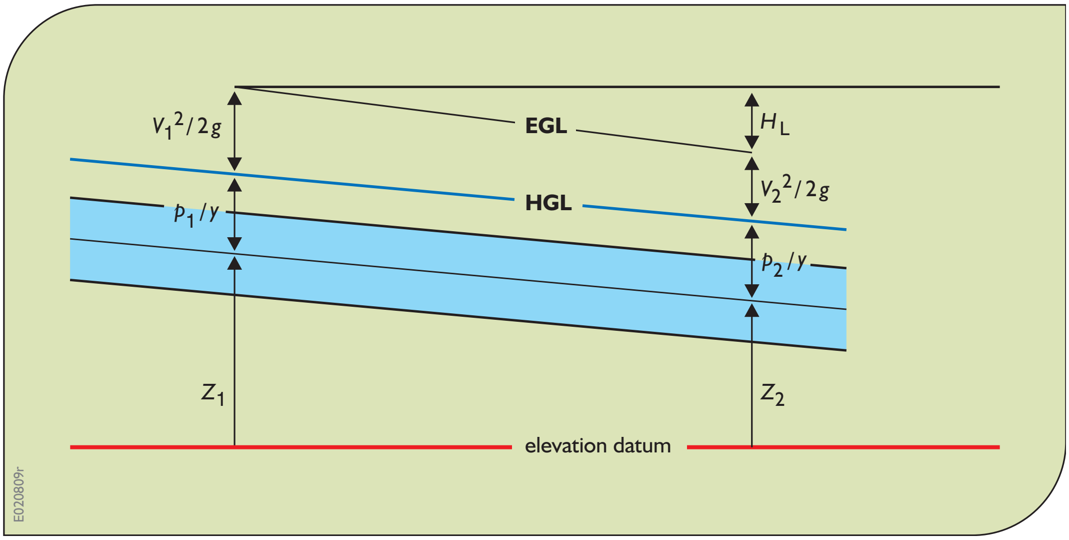



**Figure 13.5.** The energy components along a pressure pipe.

Darcy–Weisbach or Colebrook–White equations, which are used for pressure-pipe flow. They all define flow velocity, $V$ (m/s), as an empirical function of a flow resistance factor, $C$, the hydraulic radius (cross-sectional area divided by wetted perimeter), $R$ (m), and the friction or energy slope, $S = H_L$/Length.

$$V = kCR^xS^y \tag{13.2}$$

The terms $k$, $x$ and $y$ of Equation 13.2 are parameters. The roughness of the flow channel usually determines the flow resistance or roughness factor, $C$. The value of $C$ may also be a function of the channel shape, depth and fluid velocity. Values of $C$ for different types of pipes are listed in hydraulics texts or handbooks (e.g. Chin, 2000; Mays, 2000, 2005).

### 2.3.1. Open Channel Networks

For open-channel flow, Manning's or Strickler's equation is commonly used to predict the average velocity, $V$ (m/s), and the flow, $Q$ ($m^3$/s), associated with a given cross-sectional area, $A$ ($m^2$). The velocity depends on the hydraulic radius $R$ (m) and the slope $S$ of the channel as well as a friction factor $n$.

$$V = (R^{2/3}S^{1/2})/n \tag{13.3}$$

$$Q = AV \tag{13.4}$$

The values of various friction factors $n$ can be found in tables in hydraulics texts and handbooks.

The energy balance between two ends of a channel segment is defined in Equation 13.5. For open-channel flow the pressure heads are 0. Thus, for a channel containing water flowing from site $i$ to site $j$:

$$[Z + V^2/2g]_{\text{site } i} = [Z + V^2/2g + H_L]_{\text{site } j} \tag{13.5}$$

The head loss $H_L$ is assumed to be primarily due to friction.

The friction loss is computed on the basis of the average rate of friction loss along the segment and the length of the segment. This is the difference in the energy grade line elevations between sites $i$ and $j$;

$$\begin{aligned} H_L &= (EGL_1 - EGL_2) \\ &= [Z + V^2/2g]_{\text{site } i} - [Z + V^2/2g]_{\text{site } j} \end{aligned} \tag{13.6}$$

The friction loss per unit distance along the channel is the average of the friction slopes at the two ends divided by the channel length. This defines the energy grade line, $EGL$.

### 2.3.2. Pressure Pipe Networks

The Hasen–Williams equation is commonly used to predict the flows or velocities in pressure pipes. Flows and velocities are again dependent on the slope, $S$, the hydraulic radius, $R$ (m), (which equals half the pipe radius, $r$) and the cross-sectional area, $A$ ($m^2$).

$$V = 0.849\ CR^{0.63}S^{0.54} \tag{13.7}$$

$$Q = AV = \pi r^2 V \tag{13.8}$$

The head loss along a length $L$ (m) of pipe of diameter $D$ (m) containing a flow of $Q$ ($m^3/s$) is defined as

$$H_L = KQ^{1.85} \tag{13.9}$$

where $K$ is the pipe coefficient defined by Equation 13.10.

$$K = [10.66\ L]/[C^{1.85}D^{4.87}] \tag{13.10}$$

Another pipe flow equation for head loss is the Darcy–Weisbach equation based on a friction factor $f$:

$$H_L = fLV^2/D\ 2g \tag{13.11}$$

The friction factor is dependent on the Reynolds number and the pipe roughness and diameter.

Given these equations, it is possible to compute the distribution of flows and heads throughout a network of open channels or pressure pipes. The two conditions are the continuity of flows at each node, and the continuity of head losses in loops for each time period $t$.

At each node $i$:

$$Storage_{it} + Q_{it}^{\text{in}} - Q_{it}^{\text{out}} = Storage_{i,t+1} \tag{13.12}$$

In each section between nodes $i$ and $j$:

$$H_{Lit} = H_{Ljt} + H_{Lijt} \tag{13.13}$$

where the head loss between nodes $i$ and $j$ is $H_{Lijt}$.

To compute the flows and head losses at each node in Figure 13.6 requires two sets of equations, one for continuity of flows, and the other continuity of head losses. In this example, the direction of flow in two links, from $A$ to $C$, and from $B$ to $C$, are assumed unknown and hence each is represented by two non-negative flow variables.

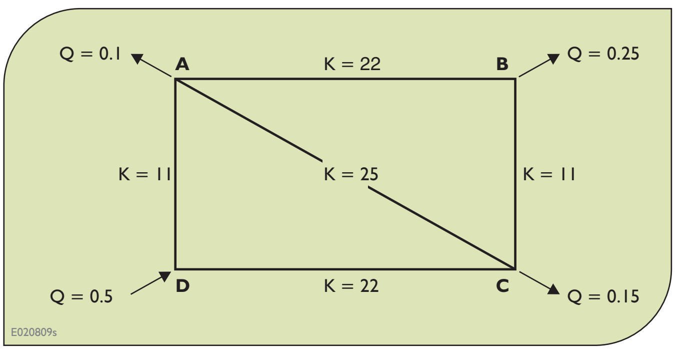



**Figure 13.6.** An example of a pipe network, showing the values of $K$ for predicting head losses from Equation 13.10.

Let $Q_{ij}$ be the flow from site $i$ to site $j$ and $H_i$ be the head at site $i$. Continuity of flow in this network requires:

$$0.5 = Q_{DA} + Q_{DC} \tag{13.14}$$

$$0.1 = Q_{DA} - Q_{AC} + Q_{CA} - Q_{AB} \tag{13.15}$$

$$0.25 = Q_{AB} + Q_{CB} - Q_{BC} \tag{13.16}$$

$$0.15 = Q_{DC} + Q_{AC} - Q_{CA} + Q_{BC} - Q_{CB} \tag{13.17}$$

Continuity of heads at each node requires:

$$H_D = H_C + 22^*(Q_{DC}^{1.85}) \tag{13.18}$$

$$H_D = H_A + 11^*(Q_{DA}^{1.85}) \tag{13.19}$$

$$H_A = H_B + 22^*(Q_{AB}^{1.85}) \tag{13.20}$$

$$H_C = H_A + 25^*((Q_{CA} - Q_{AC})^{1.85}) \tag{13.21}$$

$$H_C = H_B + 11^*((Q_{CB} - Q_{BC})^{1.85}) \tag{13.22}$$

Solving these Equations 13.14 to 13.22 simultaneously for the 5-flow and 4-head variables yields the flows $Q_{ij}$ from nodes $i$ to nodes $j$ and heads $H_i$ at nodes $i$ listed in Table 13.2. Increasing $H_D$ will increase the other heads accordingly.

The solution shown in Table 13.2 assumes no elevation heads, no storage capacity and no minor losses. Losses are usually expressed as a linear function of the velocity head, due to hydraulic structures (such as valves,

E020903k

| | | |
|---|---|---|
| $Q_{DA}$ | = | 0.29 |
| $Q_{DC}$ | = | 0.21 |
| $Q_{AC}$ | = | 0.07 |
| $Q_{CA}$ | = | 0.00 |
| $Q_{AB}$ | = | 0.12 |
| $Q_{CB}$ | = | 0.13 |
| $Q_{BC}$ | = | 0.00 |
| $H_A$ | = | 0.43 |
| $H_B$ | = | 0.00 |
| $H_C$ | = | 0.26 |
| $H_D$ | = | 1.52 |

**Table 13.2.** Flows and heads of the network shown in Figure 13.6.

restrictions or meters) at each node. This solution suggests that the pipe section between nodes *A* and *C* may not be economical, at least for these flow conditions. Other flow conditions may prove otherwise. But even if they do not, this pipe section increases the reliability of the system, and reliability is an important consideration in water supply distribution networks.

### 2.3.3. Water Quality

Many of the water quality models discussed in Chapter 12 can be used to predict water quality constituent concentrations in open channels and in pressure pipes. It is usually assumed that there is complete mixing, for example at junctions or in short segments of pipe. Reactions among constituents can occur as water travels through the system at predicted velocities. Water resident times (the ages of waters) in the various parts of the network are important variables for water quality prediction, as constituent decay, transformation and growth processes take place over time.

Computer models typically use numerical methods to find the hydraulic flow and head relationships as well as the resulting water quality concentrations. Most numerical models assume combinations of plug flow (advection) along pipe sections and complete mixing within segments of each pipe section at the end of each simulation time step. Some models also use Lagrangian approaches for tracking particles of constituents within a network. These methods are discussed in more detail in Chapter 12.

Computer programs (e.g. EPANET) exist that can perform simulations of the flows, heads and water quality behaviour within pressurized networks of pipes, pipe junctions, pumps, valves and storage tanks or reservoirs. These programs are designed to predict the movement and fate of water constituents within distribution systems. They can be used for many different kinds of application in distribution systems design, hydraulic model calibration, chlorine residual analysis and consumer exposure assessment. They can also be used to compare and evaluate the performance of alternative management strategies for improving water quality throughout a system. These can include:

- altering the sources within multiple source systems
- altering pumping and tank filling/emptying schedules
- use of satellite treatment, such as re-chlorination at storage tanks
- targeted pipe cleaning and replacement.

Computer models that simulate the hydraulic and water quality processes in water distribution networks must be run long enough for the system to reach equilibrium conditions, i.e. conditions not influenced by initial boundary assumptions. Equilibrium conditions within pipes are reached relatively quickly compared to those in storage tanks.

# 3. Wastewater

Wastewater issues include its production, its collection and its treatment prior to disposal.

## 3.1. Wastewater Production

Wastewater treatment plant influent is usually a mixture of wastewater from households and industries, urban runoff and infiltrating groundwater. The characterization of the influent, both in dry weather situations and during rainy weather, is of importance for the design and operation of the treatment facilities. In general, wastewater treatment plants can handle pure domestic wastewater better than diluted influent with low concentrations of pollutants. The discharge of urban runoff to the wastewater treatment plant dilutes the wastewater, thus affecting the treatment efficiency. The amount of infiltrating groundwater can also be significant in areas with old sewage systems.

## 3.2. Sewer Networks

Sewer flows and their pollutant concentrations vary throughout a typical day, a typical week, and over the seasons of a year. Flow conditions can range from free surface to surcharged flow, from steady to unsteady flow, and from uniform to gradually or rapidly varying non-uniform flow.

Urban drainage ditches normally have uniform cross sections along their lengths and uniform gradients. Because the dimensions of the cross sections are typically one or two orders of magnitude less than the lengths of the conduit, unsteady free-surface flow can be modelled using one-dimensional flow equations.

When modelling the hydraulics of flow it is important to distinguish between the speed of propagation of the kinematic wave disturbance and the speed of the bulk of the water. In general the wave travels faster than the water particles. Thus if water is injected with a tracer, the tracer lags behind the wave. The speed of the wave disturbance depends on the depth, width and velocity of the flow.

Flood attenuation (or subsidence) is the decrease in the peak of the wave as it propagates downstream. Gravity tends to flatten, or spread out, the wave along the channel. The magnitude of the attenuation of a flood wave depends on the peak discharge, the curvature of the wave profile at the peak, and the width of flow. Flows can be distorted (changed in shape) by the particular channel characteristics.

Additional features of concern to hydraulic modellers are the entrance and exit losses to the conduit. Typically, at each end of the conduit is an access-hole. These are storage chambers that provide access to the conduits upstream and downstream. Access-holes induce some additional head loss.

Access-holes usually cause a major part of the head losses in sewage systems. An access-hole loss represents a combination of the expansion and contraction losses. For pressure flow, the head loss, $H_L$, due to contraction can be written as a function of the downstream velocity, $V_D$, and the upstream and downstream flow cross-sectional areas $A_U$ and $A_D$:

$$H_L = K(V_D^2/2g)\ [1 - (A_D/A_U)]^2 \tag{13.23}$$

The coefficient $K$ varies between 0.5 for sudden contraction and about 0.1 for a well-designed gradual contraction.

An important parameter of a given open-channel conduit is its capacity: the flow that it can take without surcharging or flooding. Assuming normal depth flow where the hydraulic gradient is parallel to the bed of the conduit, each conduit has an upper limit to the flow that it can accept.

Pressurized flow is much more complex than free-surface flow. In marked contrast to the propagation speed of disturbances under free-surface flow conditions, the propagation of disturbances under pressurized flow in a 1 m circular conduit 100 m long can be less than a second. Some conduits can have the stable situation of free-surface flow upstream and pressurized flow downstream.

## 3.3. Wastewater Treatment

The wastewater generated by residences, businesses and industries in a community consists largely of water. It often contains less than 10% dissolved and suspended solid material. Its cloudiness is caused by suspended particles whose concentrations in untreated sewage range from 100 to 350 mg/l. One measure of the strength of the wastewater is its biochemical oxygen demand, or $BOD_5$. $BOD_5$ is the amount of dissolved oxygen aquatic microorganisms will require in five days as they metabolize the organic material in the wastewater. Untreated sewage typically has a $BOD_5$ concentration ranging from 100 mg/l to 300 mg/l.

Pathogens or disease-causing organisms are also present in sewage. Coliform bacteria are used as an indicator of disease-causing organisms. Sewage also contains nutrients (such as ammonia and phosphorus), minerals and metals. Ammonia can range from 12 to 50 mg/l and phosphorus can range from 6 to 20 mg/l in untreated sewage.

As illustrated in Figures 13.7 and 13.8, wastewater treatment is a multi-stage process. The goal is to reduce or remove organic matter, solids, nutrients, disease-causing organisms and other pollutants from wastewater before it is released into a body of water or on to the land, or is reused. The first stage of treatment is called preliminary treatment.

Preliminary treatment removes solid materials (sticks, rags, large particles, sand, gravel, toys, money, or anything people flush down toilets). Devices such as bar screens and grit chambers are used to filter the wastewater as it enters a treatment plant, and it then passes on to what is called primary treatment.

Clarifiers and septic tanks are generally used to provide primary treatment, which separates suspended solids and greases from wastewater. The wastewater is held in a tank for several hours, allowing the particles to settle to the bottom and the greases to float to the top. The solids that are drawn off the bottom and skimmed off the top receive further treatment as sludge. The clarified wastewater flows on to the next, secondary stage of wastewater treatment.

This secondary stage typically involves a biological treatment process designed to remove dissolved organic matter from wastewater. Sewage microorganisms cultivated

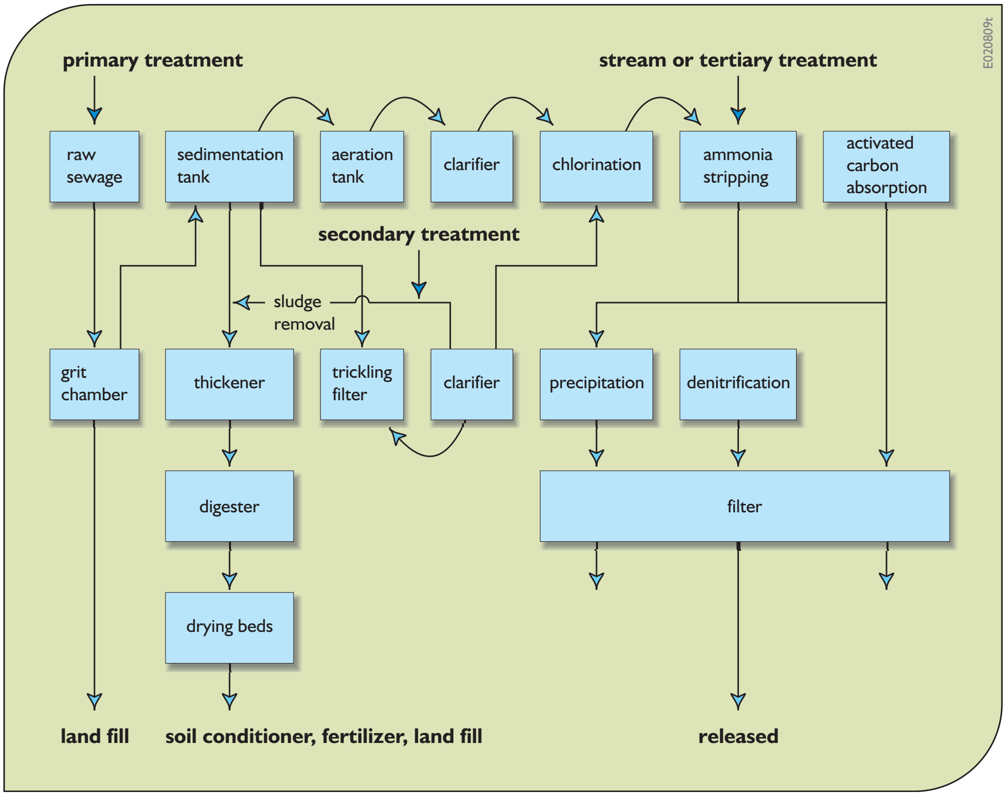



**Figure 13.7.** A typical wastewater treatment plant showing the sequence of processes for removing impurities.

**Figure 13.8.** Wastewater treatment plant in Soest, the Netherlands (Waterschap Vallei en Eem).

and added to the wastewater absorb organic matter from sewage as their food supply. Three approaches are commonly used to accomplish secondary treatment: fixed-film, suspended-film and lagoon systems.

Fixed-film systems grow microorganisms on substrates such as rocks, sand or plastic, over which the wastewater is poured. As organic matter and nutrients are absorbed from the wastewater, the film of microorganisms grows and thickens. Trickling filters, rotating biological contactors and sand filters are examples of fixed-film systems.

Suspended-film systems stir and suspend microorganisms in wastewater. As the microorganisms absorb organic matter and nutrients from the wastewater, they grow in size and number. After the microorganisms have been suspended in the wastewater for several hours, they are settled out as sludge. Some of the sludge is pumped back into the incoming wastewater to provide 'seed' microorganisms. The remainder is sent on to a sludge treatment process. Activated sludge, extended aeration, oxidation ditch and sequential batch reactor systems are all examples of suspended-film systems.

Lagoons, where used, are shallow basins that hold the wastewater for several months to allow for the natural degradation of sewage. These systems take advantage of natural aeration and microorganisms in the wastewater to renovate sewage.

Advanced treatment is necessary in some systems to remove nutrients from wastewater. Chemicals are sometimes added during the treatment process to help remove phosphorus or nitrogen. Some examples of nutrient removal systems are coagulant addition for phosphorus removal and air stripping for ammonia removal.

Final treatment focuses on removal of disease-causing organisms from wastewater. Treated wastewater can be disinfected by adding chlorine or by exposing it to sufficient ultraviolet light. High levels of chlorine may be harmful to aquatic life in receiving streams, so treatment systems often add a chlorine-neutralizing chemical to the treated wastewater before stream discharge.

Sludges are generated throughout the sewage treatment process. This sludge needs to be treated to reduce odours, remove some of the water and reduce volume, decompose some of the organic matter and kill disease-causing organisms. Following sludge treatment, liquid and cake sludges free of toxic compounds can be spread on fields, returning organic matter and nutrients to the soil.

Artificial wetlands and ponds are sometimes used for effluent polishing. In the wetlands the natural diurnal variation in the oxygen concentration is restored. Furthermore, artificial wetlands can reduce the nutrient content of the effluent by the uptake of nitrogen and phosphorus by algae or macrophytes. The organic matter may be harvested from the ponds and wetlands.

A typical model for the simulation of the treatment processes in wastewater treatment plants is the Activated Sludge Model (Gujer et al., 1999; Henze et al., 1999; Hvitved-Jacobsen et al., 1998). Activated sludge models predict the production of bacterial biomass and the subsequent conversion of organic matter and nutrients into sludge, $CO_2$ and $N_2$ gas.

# 4. Urban Drainage

Urban drainage involves:

- rainfall and surface runoff
- surface loading and washoff of pollutants
- stormwater sewer and pipe flow
- sediment transport
- structures and special flow characteristics
- separation of solids at structures
- outfalls.

These components or processes are briefly discussed in the following sub-sections.

## 4.1. Rainfall

Rainfall and the need to collect urban stormwater are the primary reasons for urban drainage systems. Storms are a major source of flow into the system. Even sanitary sewage systems that are claimed to be completely separate from storm drainage sewers are often influenced by rainfall through illicit connections or even infiltration.

Rainfall varies over time and space. These differences are normally small when considering short time periods and small distances, but they increase as time and distance increase. The ability to account for spatial differences in rainfall depends on the size of the catchment area and on the number of functioning rainfall recording points in the catchment. The use of radar permits more precision over space and time, rather as if more rain gauges were used and they were monitored more frequently. In practice, spatial effects are not measured at high resolution, and therefore events where significant spatial variations occur, such as summer thunderstorms, are usually not very accurately represented.

There are two categories of rainfall records: recorded (real) events and synthetic (not-real) events. Synthetic rainfall comes in two forms: as stochastically generated rainfall data and as design storms. The latter are derived from analyses of actual rainfall data and are used to augment or replace those historical (real) data.

Design events are a synthesized set of rainfall profiles that have been processed to produce storms with specific return periods; in other words, how often, on average, one can expect to observe rainfall events of that magnitude or greater. Design events are derived to reduce the number of runs needed to analyse system performance under design flow conditions.

### 4.1.1. Time Series Versus Design Storms

Professionals debate whether design rainfall is better represented by real rainfall or synthetic design events. The argument in favour of using synthetic storms is that they are easy to use and require only a few events to assess the system performance. The argument in favour of a time series of real rainfall is that these data include a wider range of conditions, and therefore are likely to contain the conditions that are critical in each catchment.

The two methods are not contradictory. The use of real rainfall involves some synthesis in choosing which storms to use in a time series, and in adjusting them for use on a catchment other than the one where they were measured. Time series of rainfalls are generally used to look at aspects such as overflow spill frequencies and volumes. On the other hand, synthetic design storms can be generated for a wide range of conditions including the same conditions as represented by real rainfall. This is generally considered appropriate for looking at pipe network performance.

### 4.1.2. Spatial-Temporal Distributions

Rainfall varies in space as well as in time, and the two effects are related. Short storms typically come from small rain cells that have a short life, or that move rapidly over the catchment. As these cells are small (in the order of a kilometre in diameter), there is significant spatial variation in rainfall intensity. Longer storms tend to come from large rainfall cells associated with large weather systems. These have less spatial intensity variation.

Rainfall is generally measured at specific sites using rain gauges. The recorded rainfall amount and intensity will not be the same at each site. Thus, in order to use recorded rainfall data we need some way to account for this spatial and temporal variation. The average rainfall over the catchment in any period of time can be more or less than the measured values at one or more gauges. The runoff from a portion of a catchment exposed to a high-intensity rainfall will be more than the runoff from the same amount of rainfall spread evenly over the entire catchment.

### 4.1.3. Synthetic Rainfall

A convenient way of using rainfall data is to analyse long rainfall records to define the statistical characteristics of the rainfall, and then to use these statistics to produce synthetic rainstorms of various return periods and durations.

Three parameters are used to describe the statistics of rainfall depth.

- the rainfall intensity or depth of rain in a certain period
- the length of the period over which that intensity occurs
- the frequency with which it is likely to occur, or the probability of it occurring in any particular year.

In most of the work on urban drainage and river modelling, the risks of occurrence are expressed not by probabilities but by the inverse of probability, the return period. An event that has a probability of 0.2 of being equalled or exceeded each year has an expected return period of 1/0.2 or 5 years. An event having a probability of 0.5 of being equalled or exceeded has an expected return period of 1/0.5 = 2 years.

Rainfall data show an intensity–duration–frequency relationship. The intensity and duration are inversely related. As the rainfall duration increases, the intensity reduces. The frequency and intensity are inversely related, so that as the event becomes less frequent the intensity increases.

An important part of this duration–intensity relationship is the period of time over which the intensity is averaged. It is not necessarily the length of time for which it rained from start to finish. In fact any period of rainfall can be analysed for a large range of durations, and each duration could be assigned a different return period. The largest return period might be quoted as the 'return period of the storm', but it is only meaningful when quoted with its duration.

Intensity–duration–frequency relationships or depth–duration–frequency relationships, as shown in Figure 13.9, are derived by analysis of a long set of rainfall records. Intensity–duration–frequency data are commonly available all over the world and therefore it is important to be aware of how they are derived and ways they can be used for simulation modelling.

The depth of rainfall is the intensity times its duration integrated over the total storm duration.

### 4.1.4. Design Rainfall

Design rainfall events (hyetographs) for use in simulation models are derived from intensity–duration–frequency data.

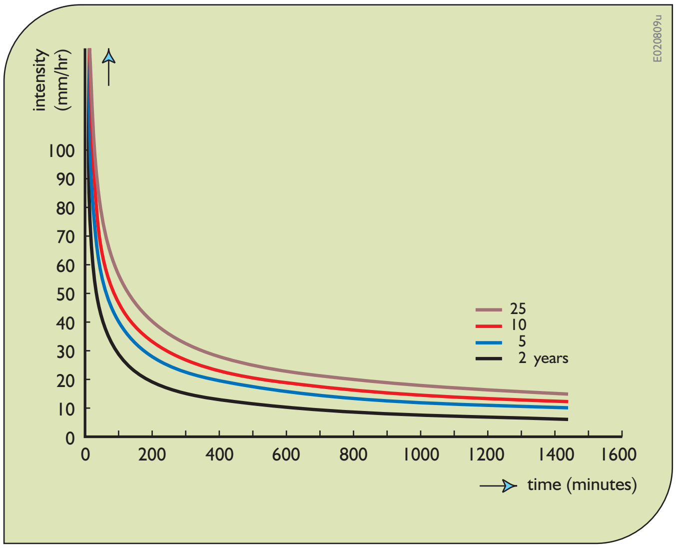



**Figure 13.9. Rainfall intensity–duration–frequency (return period) curves.**

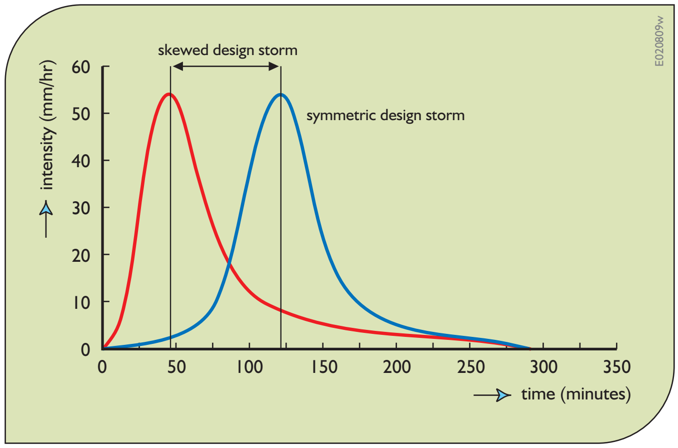



**Figure 13.10. Storm peak skewness profiles.**

The rainfall intensity during an event is not uniform in time, and its variation both in intensity and in the time when the peak intensity occurs during the storm can be characterized by the peakedness of the storm and the skew of the storm (Figure 13.10).

A design storm is a synthetic storm that has an appropriate peak intensity and storm profile.

## 4.2. Runoff

Runoff prediction is often the first step in obtaining stream and river flow data needed for various analyses. Often lack of detailed site-specific land cover, soil conditions and precipitation data as well as runoff processes limit the accuracy of runoff predictions.

### 4.2.1. Runoff Modelling

The runoff from rainfall involves a number of processes and events, as illustrated in Figure 13.11, and can be modelled using various methods. Most of these methods assume an initial loss, a continuing loss, and a remainder contributing to the system runoff.

Most models assume that the first part of a rainfall event goes to initial wetting of surfaces and filling of depression storage. The depth assumed to be lost is usually related to the surface type and condition. Rainwater can be intercepted by vegetation or can be trapped in depressions on the ground surface. It then either infiltrates into the ground and/or evaporates. Depression storage can occur on any surface, paved or otherwise.

Initial-loss depths are defined as the minimum quantity of rainfall causing overland runoff. The initial-loss depth of rainfall for catchment surfaces can be estimated as the intercept on the rainfall axis of plots of rainfall versus runoff (Figure 13.12). The runoff values shown in Figure 13.12 were obtained for various catchments in the United Kingdom (Price, 2002).

As rainfall increases, so does depression storage. The relationship between depression storage and surface slope $S$ is assumed to be of the form $aS^{-b}$, where $S$ is average slope of the sub-catchment and $a$ and $b$ are parameters between 0 and 1. The values of $a$ and $b$ depend in part on the surface type.

Evaporation, another source of initial loss, is generally considered to be relatively unimportant. For example, in the case of a heavy summer storm (25-mm rainfall depth) falling on hot asphalt (temperature say 60 °C falling to 20–30 °C as a result of sensible heat loss), a maximum evaporation loss of 1 mm is likely to occur.

Continuing losses are often separated into two parts: evapotranspiration and infiltration. These processes are usually assumed to continue throughout and beyond the storm event as long as water is available on the surface of the ground. Losses due to plant transpiration and general evaporation are not particularly an issue for single events, but can be during the inter-event periods where catchment drying takes place. This is applicable to models where time-series data are used and generated.

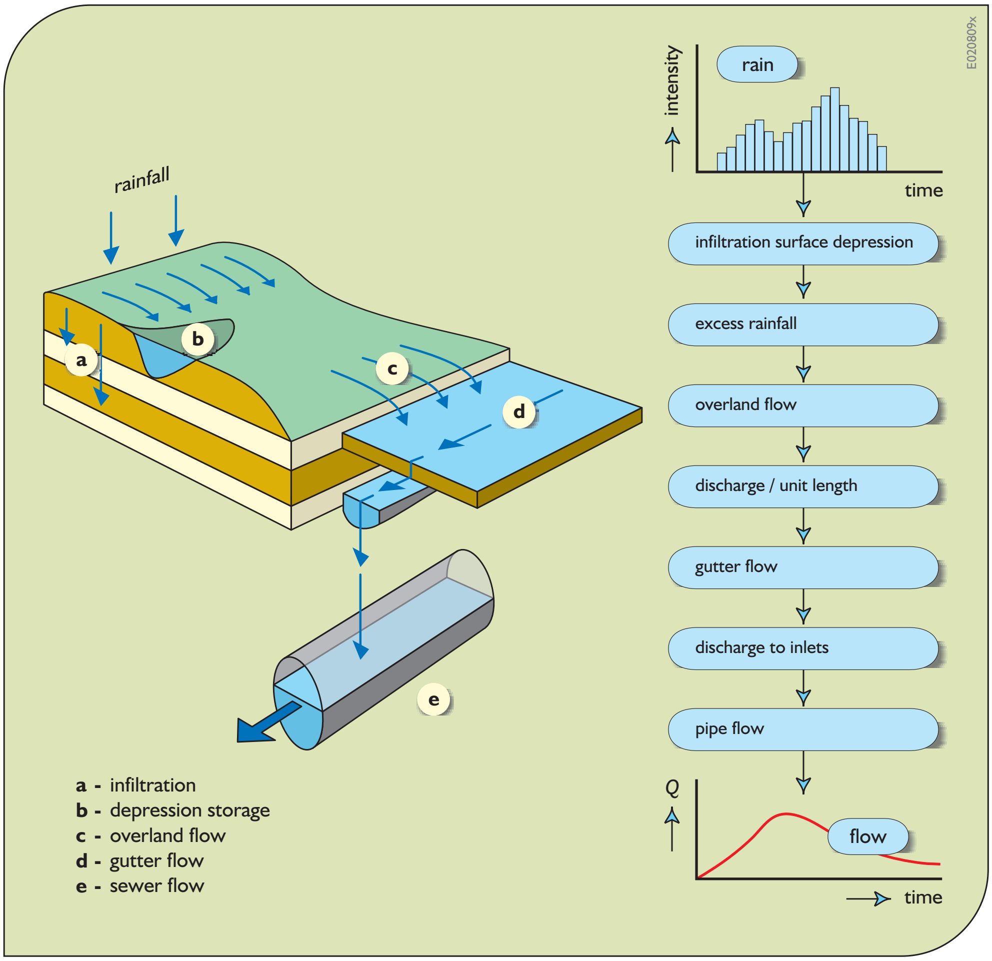



**Figure 13.11.** Schematic representation of urban rainfall–runoff processes.

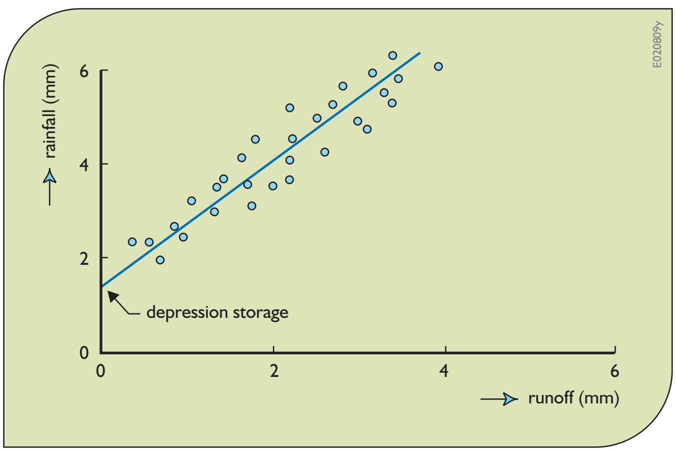



**Figure 13.12.** Estimation of depression storage based on data from catchments in the UK (Price, 2002).

Infiltration is usually assumed to account for the remaining rainfall that does not enter into the drainage system. The proportion of this loss can range from 100% for very permeable surfaces to 0% for completely impermeable surfaces.

Many models try to account for the wetting of the catchment and the increasing runoff that takes place as wetting increases. The effect of this is shown in Figure 13.13.

It is often impractical to take full account of the variability in urban topography and surface conditions. Fortunately for modellers, impervious (paved) surfaces are often dominant in an urban catchment, and the loss of rainfall prior to runoff is usually relatively small in periods with much rainfall.

Runoff-routing is the process of passing rainfall across the surface to enter the drainage network. This process results in attenuation and delay. These parameters are modelled by means of routing techniques that generally consider catchment area size, ground slope and rainfall intensity in determining the flow rate into the network.

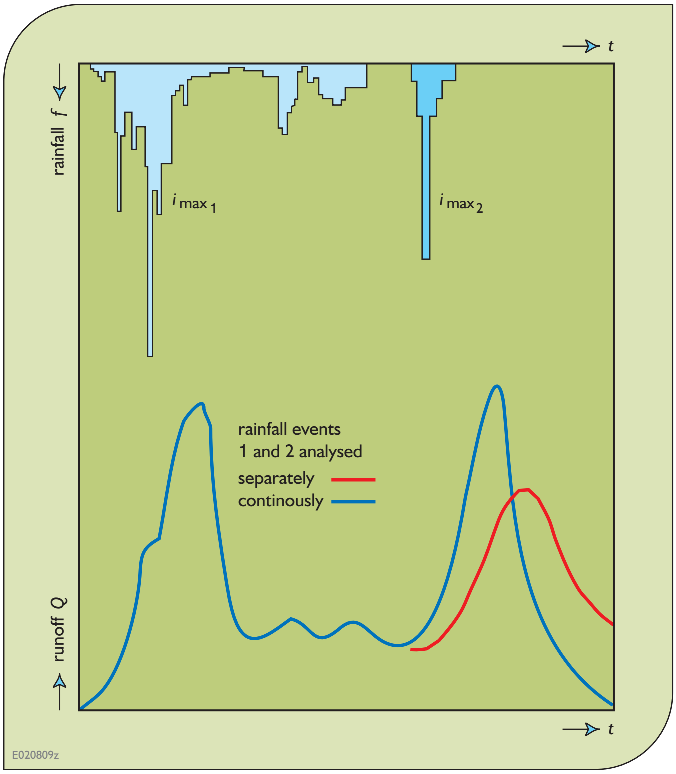



**Figure 13.13. Effect of catchment wetness on runoff *Q* over time *t* (Price, 2002).**

The topography and surface channels, and even upstream parts of the sewage system, are usually lumped together in this process and are not explicitly described in a model. The runoff-routing process is often linked to catchment surface type, and empirical calibration factors are used accordingly.

A number of models for rainfall–runoff and runoff-routing are available and are used in various parts of the world. Overland runoff on catchment surfaces can be represented by the kinematic wave equation. However, direct solution of this equation in combination with the continuity equation has not been a practical approach when applied to basins with a large number of contributing sub-catchments. Simpler reservoir-based models, which are less computationally and data demanding, represent the physical processes almost as accurately as the more complex physically based approaches (Price, 2002). In practice, models applied to catchments typically assume an average or combined behaviour of a number of overland flow surfaces, gutters and feeder pipes. Therefore, the parameters of a physically based approach (for example the roughness value) as applied would not relate directly to parameters representative of individual surfaces and structures.

Many overland flow-routing models are based on a linear reservoir-routing concept. A single reservoir model assumes the outflow, $Q(t)$ ($m^3/s$), at the catchment outlet is proportional to the volume of stormwater, $S(t)$ ($m^3$), present on the ground surface of that catchment, including the non-explicitly modelled network that contributes stormwater to that outlet point of the urban drainage system. To take into account the effects of depression storage and other initial losses, the first millimetre(s) of rainfall may not contribute to the runoff.

The basic equation for runoff $Q(t)$ ($m^3/s$) at time $t$ is:

$$Q(t) = S(t)\,K \tag{13.24}$$

where $K$ is a linear reservoir coefficient. This coefficient is sometimes a function of the catchment slope, area, length of longest sewer and rainfall intensity. For a two linear-reservoir model, two reservoirs are applied in series for each surface type with an equivalent storage – output relationship, as defined by Equation 13.24, for each reservoir.

The simplest models rely on fixed runoff coefficients $K$. They best apply to impervious areas where antecedent soil moisture conditions are not a factor.

Typical values for runoff coefficients are given in Table 13.3 (Price, 2002). Use of these coefficients should be supported either by field observations or by expert judgement.

### 4.2.2. The Horton Infiltration Model

The Horton model describes the increasing runoff from permeable surfaces while a rainfall event occurs by keeping track of decreasing infiltration as the soil moisture content increases. The runoff from paved surfaces is assumed to be constant while the runoff from permeable surfaces is a function of the conceptual wetting and infiltration processes.

On the basis of infiltrometer studies on small catchments, Horton defined the infiltration rate, $f$, either on pervious surfaces or on semi-pervious surfaces, as a function of time, $t$ (hours), the initial infiltration rate, $f_o$ (mm/hr), the minimum (limiting or critical) infiltration rate, $f_c$ (mm/hr), and an infiltration rate constant, $k$ (1/hr).

$$f = f_c + (f_o - f_c)e^{-kt} \tag{13.25}$$

**Table 13.3.** Typical values of the runoff fraction (coefficient *K*).

| surface type | description | coefficient K |
|---|---|---|
| paved | high quality paved roads with gullies < 100m apart | 1.00 |
| paved | high quality paved roads with gullies > 100m apart | 0.90 |
| paved | medium quality paved roads | 0.85 |
| paved | poor quality paved roads | 0.80 |
| permeable | high to medium density housing | 0.55 - 0.45 |
| permeable | low density housing or industrial areas | 0.35 |
| permeable | open areas | 0.00 - 0.25 |

E020903m

The minimum or limiting infiltration rate, $f_c$, is commonly set to the saturated groundwater hydraulic conductivity for the applicable soil type.

The integration of Equation 13.25 over time defines the cumulative infiltration $F(t)$:

$$F(t) = f_c t + (f_o - f_c)(1 - e^{-kt})/k \tag{13.26}$$

The Horton equation variant as defined in Equation 13.26 represents the potential infiltration depth, $F$, as a function of time, $t$, assuming the rainfall rate is not limiting, i.e. it is higher than the potential infiltration rate. Expressed as a function of time, it is not suited for use in a continuous simulation model. The infiltration capacity should be reduced in proportion to the cumulative infiltration volume, $F$, rather than in proportion to time. To do this, Equation 13.26 may be solved iteratively to find the time it takes to cause ponding, $t_p$, as a function of $F$. That ponding time $t_p$ is used in Equation 13.25 to establish the appropriate infiltration rate for the next time interval (Bedient and Huber, 1992). This procedure is used, for example, in the urban stormwater management model (SWMM) (Huber and Dickinson, 1988).

A flow chart of the calculations performed in a simulation program in which the rainfall can vary might be as shown in Figure 13.14.

Various values for Horton's infiltration model are available in the published literature. Values of $f_o$ and $f_c$, as determined by infiltrometer studies, Table 13.4, are highly variable, even by an order of magnitude on seemingly similar soil types. Furthermore, the direct transfer of values as measured on rural catchments to urban catchments is not advised due to the compaction and vegetation differences associated with the latter surfaces.

### 4.2.3. The US Soil Conservation Method (SCS) Model

The SCS model (USDA, 1972) is widely used, especially in the United States, France, Germany, Australia and parts of Africa, for predicting runoff from rural catchments. It has also been used for the permeable component in a semi-urban environment. This runoff model allows for variation in runoff depending on catchment wetness. The model relies on what are called curve numbers, $CN$.

The basis of the method is the continuity equation. The total depth (mm) of rainfall, $R$, either evaporates or is otherwise lost, $I_a$, infiltrates and is retained in the soil, $F$, or runs off the land surface, $Q$. Thus,

$$R = I_a + F + Q \tag{13.27}$$

The relationship between the depths (mm) of rainfall, $R$, runoff, $Q$, the actual retention, $F$, and the maximum potential retention storage, $S$ (not including $I_a$), is assumed to be

$$F/S = Q/(R - I_a) \tag{13.28}$$

when $R > I_a$. These equations combine to give the SCS model

$$Q = (R - I_a)^2/(R - I_a + S) \tag{13.29}$$

This model can be modified for use in continuous simulation models.

Numerical representation of the derivative of Equation 13.29 can be written as Equation 13.30 for predicting

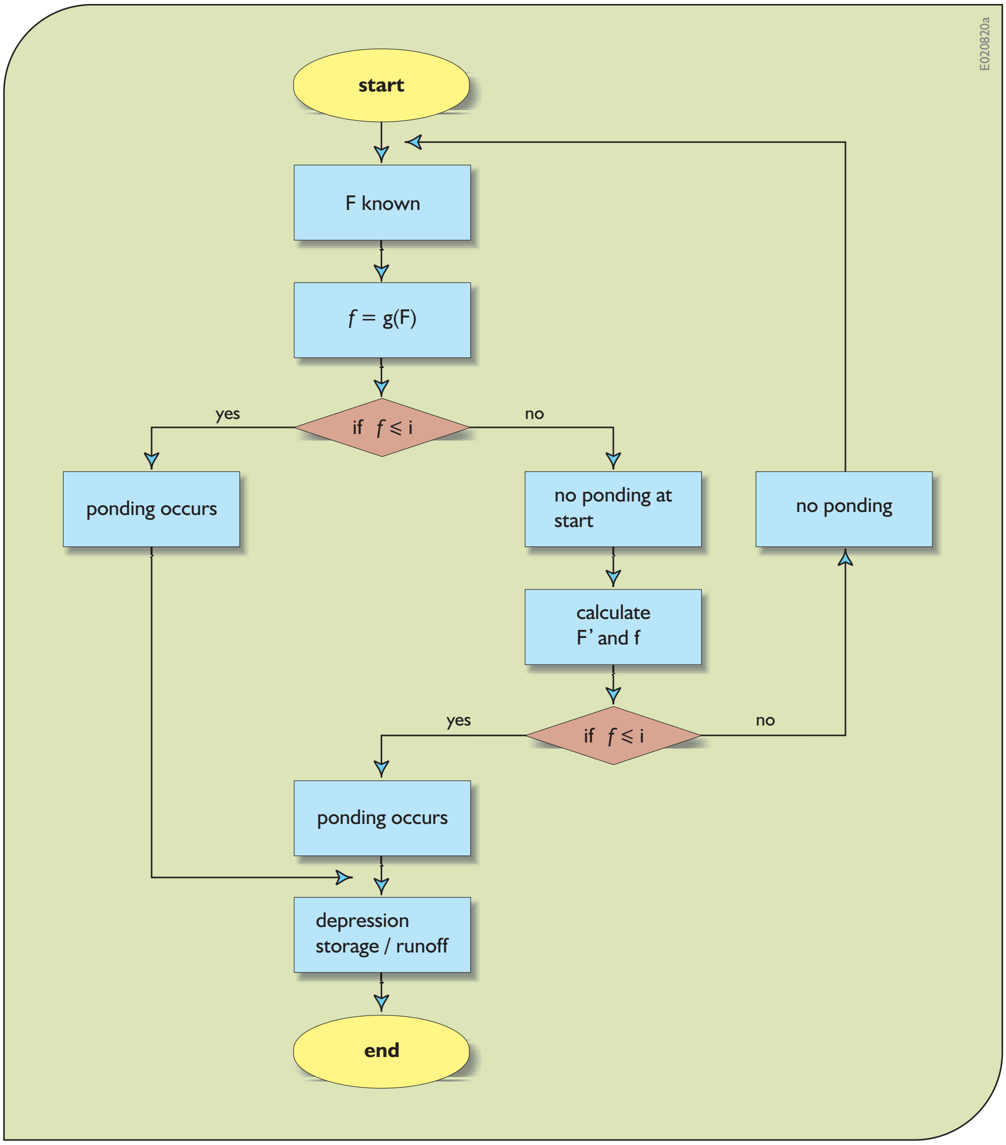



**Figure 13.14.** Flow chart of Horton model infiltration algorithm used in each time step of a simulation model.

E020903n

| SCS soil group | $f_o$ (mm/hr) | $f_c$ (mm/hr) | $k$ (1 / hr) |
|---|---|---|---|
| **A** | 250 | 25.4 | 2 |
| **B** | 200 | 12.7 | 2 |
| **C** | 125 | 6.3 | 2 |
| **D** | 76 | 2.5 | 2 |

**Table 13.4.** Values for Horton's infiltration model for different soil groups as defined by the Soil Conservation Service. Note: see Table 13.6 for SCS soil groups.

the runoff, $q$ (mm/$\Delta t$), over a time interval $\Delta t$ given the rainfall $r$ (mm/$\Delta t$), in that time interval.

$$q = r(R - I_a)(R - I_a + 2S)/(R - I_a + S)^2 \qquad (13.30)$$

This equation is used incrementally, enabling the rainfall and runoff coefficients, $r$ and $q$, to change during the event.

The two parameters $S$ and $I_a$ are assumed to be linearly related by

$$I_a = kS \qquad (13.31)$$

where $0 < k < 0.2$.

The original SCS approach recommended $k = 0.2$. However other studies suggest $k$ values between 0.05 and 0.1 may be more appropriate.

The storage variable, $S$, is itself related to an index known as the runoff curve number, $CN$, representing the combined influence of soil type, land management practices, vegetation cover, urban development and antecedent moisture conditions on hydrological response. $CN$ values vary between 0 and 100, 0 representing no runoff and 100 representing 100% runoff.

The storage parameter $S$ is related to the curve number $CN$ by

$$S = (25400/CN) - 254 \qquad (13.32)$$

Curve number values depend on antecedent moisture condition classes (AMC) and hydrological soil group. The antecedent moisture conditions are divided into three classes, as defined in Table 13.5.

The four hydrological soil groups are defined in Table 13.6.

The $CN$ value can either be defined globally for the catchment model or can be associated with specific surface types. $CN$ values for different conditions are available from various sources. Table 13.7 lists some of those relevant to urban areas and antecedent moisture condition class AMC II.

E020903o

| AMC | total 5-day antecedent rainfall (mm), season: dormant | growing |
|---|---|---|
| I | < 12.5 | < 35.5 |
| II | 12.5 - 28.0 | 35.5 - 53.5 |
| III | > 28.0 | > 53.5 |

**Table 13.5.** Antecedent moisture condition classes (AMC) for determining curve numbers *CN.*

Figure 13.15 identifies the $CN$ values for antecedent moisture condition classes (AMC) I and III based on class II values.

### 4.2.4. Other Rainfall–Runoff Models

The rainfall–runoff element of the popular SWMM model generally assumes 100% runoff from impermeable surfaces and uses the Horton or the Green–Ampt model

**Table 13.6.** SCS hydrological soil groups used in Tables 13.4 and 13.7.

| soil type | definition |
|---|---|
| A | (low runoff potential)<br>high infiltration rates even when thoroughly wetted.<br>chiefly deep, well to excessively drained sands or gravels.<br>high rate of water transmission. |
| B | moderate infiltration rates when thoroughly wetted.<br>chiefly moderately deep to deep, moderately-well to well drained soils with moderately-fine to moderately-coarse textures.<br>moderate rate of water transmission. |
| C | slow infiltration rates when thoroughly wetted.<br>chiefly solids with layer that impedes downward movement of water, or soils with a moderately-fine to fine texture.sslow rate of water transmission. |
| D | (high runoff potential)<br>very slow infiltration and transmission rates when thoroughly wetted.<br>chiefly:<br>• clay soils with a high swelling potential<br>• soils with a permanent high water table<br>• soils with a clay pan or clay layer at or near the surface<br>• shallow soils over nearly impervious material |

E020903p

| land cover class | land cover / landuse / treatment | stormflow potential | hydrological soil group A | B | C | D |
|---|---|---|---|---|---|---|
| urban & sub-urban land use | open spaces, parks, cemeteries | 75 % grass | 39 | 61 | 74 | 80 |
| | open spaces, parks, cemeteries | 75 % grass | 49 | 69 | 79 | 84 |
| | **commercial / business area** | **% area impervious** | **A** | **B** | **C** | **D** |
| | industrial districts | 85 % | 89 | 92 | 94 | 95 |
| | residential: lot size 500 m² | 72 % | 81 | 88 | 91 | 93 |
| | 1000 m² | 65 % | 77 | 85 | 90 | 92 |
| | 1350 m² | 38 % | 61 | 75 | 83 | 87 |
| | 2000 m² | 30 % | 57 | 72 | 81 | 86 |
| | 4000 m² | 25 % | 54 | 70 | 80 | 85 |
| | paved parking lots, roofs, etc. | 20 % | 51 | 68 | 78 | 84 |
| | streets/roads: tarred, with storm sewers, curbs | | 98 | 98 | 98 | 98 |
| | gravel | | 76 | 85 | 89 | 91 |
| | dirt | | 72 | 82 | 87 | 89 |
| | dirt-hard surface | | 74 | 84 | 90 | 92 |

E020903q

**Table 13.7.** Initial *CN* values for AMC II with various urban land use, cover quality and hydrological soil groups (defined in Table 13.6).

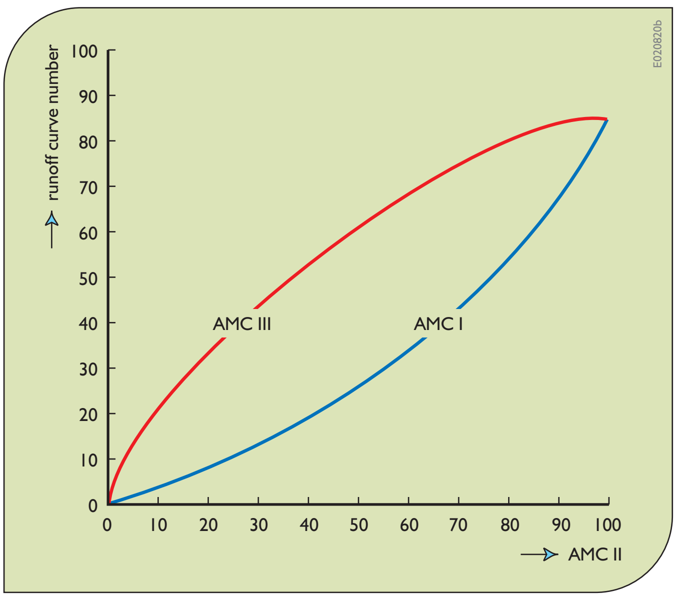



**Figure 13.15.** Runoff curve numbers for AMC classes I and III based on curve numbers for AMC II.

for permeable runoff. The Green–Ampt model is similar to the Horton model in that it has a conceptual infiltration rate that varies with time. It is therefore applicable to pervious or semi-pervious catchments (Huber and Dickinson, 1988; Roesner et.al., 1988).

## 4.3. Surface Pollutant Loading and Washoff

The modelling of surface pollutant loading and washoff into sewage systems is very imprecise. Pollutants that build up on the surface of an urban area originate from wind-blown dust, debris that is both natural and human-made, including vehicle emissions. When rainfall takes place, some of this material is washed into the stormwater sewers or gullies as dissolved pollutants and fine solids. During buildup time, many of the pollutants degrade.

Deposition of this material is not homogeneous but rather is a function of climate, geography, land use and human activity. The mechanism of washoff is obviously a function of location, rainfall intensity, slope, flow rate, vehicle disturbance and so on. None of these factors is explicitly modelled in most washoff and sewer flow models.

Measurements made of pollutant accumulation and washoff have been the basis of empirical equations representing both loading and washoff processes. In practice the level of information available and the complexity of the processes being represented make the models of pollutant loading and washoff a tool whose outputs must be viewed for what they are – merely guesses.

### 4.3.1. Surface Loading

Pollutant loadings and accumulation on the surface of an urban catchment occur during dry periods between rainstorms.

A common hypothesis for pollutant accumulation during dry periods is that the mass loading rate, $m_P$ (kg/ha/day), of pollutant $P$ is constant. This assumed constant loading rate on the surface of the ground can vary over space and is related to the land use of that catchment. In reality, these loadings on the land surface will not be the same, either over space or over time. Hence, to be more statistically precise, a time series of loadings may be created from one or more probability distributions of observed loadings. (Just how this may be done is discussed in Chapter 7.) Different probability distributions may apply when, for example, weekend loadings differ from workday loadings. However, given all the other uncertain assumptions in any urban loading and washoff model, the effort may not be justified.

As masses of pollutants accumulate over a dry period they may degrade as well. The time rate of degradation of a pollutant $P$ is commonly assumed to be proportional to its total accumulated mass $M_P$ (kg/ha). Assuming a proportionality constant (decay rate constant) of $k_P$ (1/day), the rate of change in the accumulated mass $M_P$ over time $t$ is

$$dM_P/dt = m_P - k_P M_P \tag{13.33}$$

As the number of days during a dry period becomes very large, the limiting accumulation of a mass $M_P$ of pollutant $P$ is $m_P/k_P$. If there is no decay, then of course $k_P$ is 0 and the limiting accumulation is infinite.

Integrating Equation 13.33 over the duration $\Delta t$ (days) of a dry period yields the mass, $M_P(\Delta t)$ (kg/ha), of each pollutant available for washoff at the beginning of a rainstorm.

$$M_P(\Delta t) = M_P(0)e^{-kp\Delta t} + [m_P(1 - e^{-kp\Delta t})/k_P] \tag{13.34}$$

where $M_P(0)$ is the initial mass of pollutant $P$ on the catchment surface at the beginning of the dry period (that is, at the end of the previous rainstorm).

Sediments (which become suspended solids in the runoff) are among the pollutants that accumulate on the surface of urban catchments. They are important in themselves, but also because some of the other pollutants that accumulate become attached to them. Sediments are typically defined by their medium diameter size value ($d_{50}$). Normally a minimum of two sediment fractions are modelled: one coarse, high-density material (grit) and one fine (organics).

The sediments of each diameter size class are commonly assumed to have a fixed amount of pollutants attached to them. The fraction of each attached pollutant, sometimes referred to as the potency factor of the pollutant, is expressed as kg of pollutant per kg of sediment. Potency factors are one method for defining pollutant inputs into the system.

### 4.3.2. Surface Washoff

Pollutants in the washoff may be dissolved in the water, or be attached to the sediments. Many models of the transport of dissolved and particulate pollutants through a sewerage system assume each pollutant is conservative (that is, it does not degrade over time). For practical purposes this is a reasonable assumption when the time of flow in the sewers is relative short. Otherwise it may not be a good assumption, but at least it is a conservative one.

Pollutants can enter the sewage system from a number of sources. A major source is the washoff of pollutants from the catchment surface during a rainfall event. Their removal is caused by the impact of rain and by erosion from runoff flowing across the surface. Figure 13.16 shows schematically some sources of pollution in the washoff model.

The rate of pollutant washoff depends on an erosion coefficient, $\alpha_P$, and the quantity, $M_P$, of available pollutant, $P$. As the storm event proceeds and pollutants

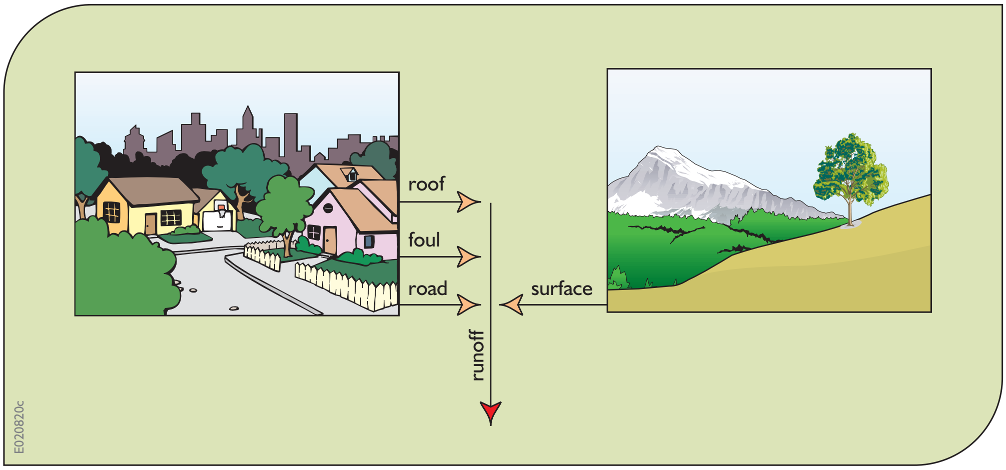



**Figure 13.16.** Some sources of pollutants in the washoff to sewage systems from land surfaces.

are removed from the catchment, the quantities of available pollutants decrease, hence the rate of pollutant washoff decreases even with the same runoff.

When runoff occurs, a fraction of the accumulated load may be contained in it. This fraction will depend on the extent of runoff. If a part of the surface loading of a pollutant is attached to sediments, its runoff will depend on the amount of sediment runoff, which in turn is dependent on the amount of surface water runoff.

The fraction of total surface loading mass contained in the runoff will depend on the runoff intensity. The following approximate relation may apply for the fraction, $f_t^R$, of pollutant $P$ in the runoff $R_t$ in period $t$:

$$f_t^P = \alpha_P R_t/(1 + \alpha_P R_t) \qquad (13.35)$$

The greater the runoff $R_t$ the greater will be the fraction $f_t^P$ of the total remaining pollutant loading in that runoff. The values of the parameters $\alpha_P$ are indicators of the effectiveness of the runoff in picking up and transporting the particular pollutant mass. Their values are dependent on the type of pollutant $P$ and on the land cover and topography of particular basin or drainage area. They can be determined on the basis of measured pollutant mass surface loadings and on the mass of pollutants contained in the rainfall and sediment runoff, preferably at the basin of interest. Since such data are difficult, or at least expensive, to obtain, they are usually based on experiments in laboratories.

A mass balance of pollutant loadings can define the total accumulated load, $M_{Pt+1}$, at the end of each simulation time period $t$ or equivalently at the beginning of each time period $t+1$. Assuming a daily simulation time step,

$$M_{Pt+1} = (1 - f_t^P)\, M_{Pt}\, e^{-kP} + m_P \qquad (13.36)$$

Of interest, of course, is the total pollutant mass in the runoff. For each pollutant type $P$ in each period $t$, these will be $f_t^P(M_{Pt})$ for the dissolved part. The total mass of pollutant $P$ in the runoff must also include those attached fractions (potency factors), if any, of each sediment size class being modelled as well.

As the sediments are routed through the system, those from different sources are mixed together. The concentrations of associated pollutants therefore change during the simulation as different proportions of sediment from different sources are mixed together. The results are given as concentrations of sediment, concentrations of dissolved pollutants and concentrations of pollutants associated with each sediment fraction.

### 4.3.3. Stormwater Sewer and Pipe Flow

Flows in pipes and sewers have been analysed extensively and their representation in models is generally accurately defined. The hydraulic characteristics of sewage are essentially the same as clean water. Time-dependent effects are, in part, a function of the change in storage in access-holes. Difficulties in obtaining convergence occur at pipes with steep to flat transitions, dry pipes and the like, and additional features and checks are therefore needed to achieve satisfactory model results.

### 4.3.4. Sediment Transport

Pollutant transport modelling of both sediment and dissolved fractions involves defining the processes of erosion and deposition, and advection and possibly dispersion. One-dimensional models by their very nature cannot predict the sediment gradient in the water column. In addition, the concept of the sewer being a bio-reactor is not included in most simulation models. Most models assume pollutants are conservative while in residence in the drainage system before being discharged into a water body. All the processes that take place in transit are generally either ignored or approximated using a range of assumptions.

### 4.3.5. Structures and Special Flow Characteristics

Access-holes, valves, pipes, pumping stations, overflow weirs and other elements that affect the flows and head losses in sewers can be explicitly included in deterministic simulation models. The impact of some of these structures can only be predicted using two- or three-dimensional models. However, the ever-increasing power of computers is making higher-dimensional fluid dynamic analyses increasingly available to practising engineers. The greatest limitation may be more related to data and calibration than to computer models and costs.

## 4.4. Water Quality Impacts

The quality of water in urban drainage and sewer systems can impact the flow conditions in those systems as well as the quality of the water into which these wastewater and drainage flows are discharged.

### 4.4.1. Slime

Slime can build up on the perimeters of sewers that contain domestic sewage. The buildup of slime may have a significant effect on roughness. In a combined system the effect will be less, as the maximum daily flow of domestic sewage will not usually be a significant part of pipe capacity.

The extent to which the roughness is increased by sliming depends on the relation between the sewage discharge and the pipe capacity. Sliming will occur over the whole of the perimeter below the water level that corresponds to the maximum daily flow. The slime growth will be heaviest in the region of the maximum water level. Over the lower part of the perimeter, the surface will still be slimed, but to a lesser extent than at the waterline. Above the maximum waterline the sewer surface will tend to be fairly free of slime.

### 4.4.2. Sediment

When sediment is present in the sewer, the roughness increases quite significantly. It is difficult to relate the roughness to the nature and time-history of the sediment deposits. Most stormwater sewers contain some sediment deposits, even if only temporarily. The only data available suggest that the increase in head loss can range from 30 to 300 mm, depending on the configuration of the deposit and on the flow conditions. The higher roughness value is more appropriate when the sewer is part full and when considerable energy is lost as a result of the generation of surface disturbances. In practice lower roughness values are used for flow states representing extreme events when sewers are operating in surcharge.

### 4.4.3. Pollution Impact on the Environment

The effects of combined sewer overflows (*CSO*s) or discharges are particularly difficult to quantify and regulate because of their intermittent and varied nature. Their immediate impact can only be measured during a spill event, and their chronic effects are often difficult to isolate from other pollution inputs. Yet they are among the major causes of poor river water quality. Standards and performance criteria specifically for intermittent discharges are therefore needed to reduce the pollutants in *CSO*s.

Drainage discharges that affect water quality can be divided into four groups:

- Those that contain oxygen-demanding substances. These can be either organic, such as fecal matter, or inorganic. (Heated discharges, such as cooling waters, reduce the saturated concentration of dissolved oxygen.)
- Those which contain substances that physically hinder re-oxygenation at the water surface, such as oils.
- Discharges containing toxic compounds, including ammonia, pesticides and some industrial effluents.
- Discharges that are high in suspended solids and thus inhibit biological activity by excluding light from the water or by blanketing the bed.

Problems arise when pollutant loads exceed the self-purification capacity of the receiving water, harming aquatic life and restricting the use of the water for consumption and for many industrial and recreational purposes. The assimilative capacity for many toxic substances is very low. Water polluted by drainage discharges can create nuisances such as unpleasant odours. It can also be a direct hazard to health, particularly in tropical regions where water-borne diseases such as cholera and typhoid are prevalent.

The aim of good drainage design, with respect to pollution, is to balance the effects of continuous and intermittent discharges against the assimilation capacity of the water, so as to achieve in a cost-effective way the desired quality of the receiving water.

Figure 13.17 shows the effect of a discharge that contains suspended solids and organic matter. The important indicators showing the effect of the discharge are the dissolved oxygen in diagram 'a' and the clean water fauna shown in diagram 'd'. The closeness with which the clean water fauna follow the dissolved oxygen reflects the reliance of a diverse fauna population on dissolved oxygen. These relationships are used by biologists to argue for greater emphasis on biological indicators of pollution, as these respond to intermittent discharges better than chemical tests which, if not continuous, may miss the pollution incident. There are a number of biological indexes in use in most countries in Europe.

In Figure 13.17 the *BOD* in diagram 'a' rises or is constant after release despite some of the pollutant being digested and depleting the dissolved oxygen. This is because there is a time lag of up to several days while the bacteria, which digest the pollutant, multiply. The suspended solids (*SS*) settle relatively quickly and they

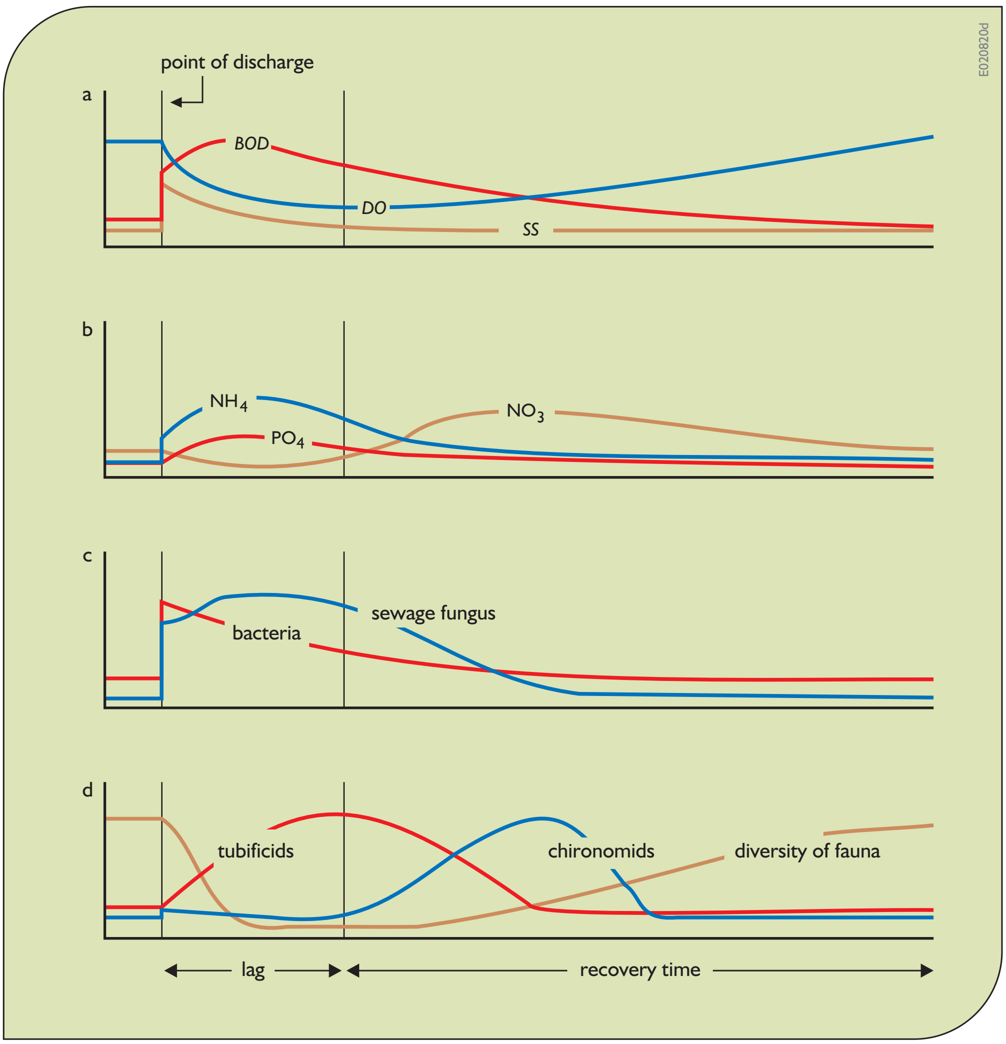



**Figure 13.17.** Pollution impact along a waterway downstream from its discharge.

can then be a source of pollutants if the bed is disturbed by high flows. This can create a subsequent pollution incident, especially if the suspended solids contain quantities of toxic heavy metals.

Diagram 'b' of Figure 13.17 shows ammonium ions ($NH_4^+$) that are discharged as part of the dissolved pollutants being oxidized to nitrates ($NO_3^-$). The rise in the ammonium concentration downstream of the discharge is relevant because of the very low tolerance many aquatic organisms, particularly fish, have to the chemical. The ammonium concentration rises if the conditions are anaerobic, and will decline once aerobic conditions return and the ammonium ions are oxidized to nitrates.

Diagrams 'c' and 'd' show the effect of combined sewer overflows on flora and fauna. The increased quantities of phosphate and nitrate nutrients that they consume can lead to eutrophication. The fauna show perhaps the clearest pattern of response. The predictability of this response has led to the development of the many biological indices of pollution. The rapid succession of organisms illustrates the pattern of dominance of only a few species in polluted conditions. For example, tubificid worms can exist in near-anaerobic conditions and, having few competitors, they can multiply prolifically. As the oxygen levels increase these organisms are succeeded by chironomids (midge larvae) and so on until in clean well-oxygenated water, there is a wide diversity of species all competing with each other.

In most circumstances the concentration of dissolved oxygen (*DO*) is the best indicator of the 'health' of a water source. A clean water source with little or no content of biodegradable pollutants will have an oxygen concentration of about 9–10 mg/l when in equilibrium with air in a temperate environment. This maximum saturation concentration is temperature dependent. The hotter the water, the lower the *DO* saturation concentration.

All higher forms of life in a river require oxygen. In the absence of toxic impurities there is a close correlation between *DO* and biodiversity. For example most game-fish die when the *DO* concentration falls below about 4 mg/l.

Perhaps of more pragmatic significance is the fact that oxygen is needed in the many natural treatment processes performed by microorganisms that live in natural water bodies. The quantity of oxygen required by these organisms to break down a given quantity of organic waste is, as previously discussed, the biochemical oxygen demand (*BOD*). It is expressed as mg of dissolved oxygen required by organisms to digest and stabilize the waste in a litre of water (mg/l). These organisms take time to fully digest and stabilize the waste. The rate at which they do so depends on the temperature of the water at the start, and the quantity of these organisms available in it.

Since the *BOD* test measures only the biodegradable material, it may not give an accurate assessment of the total quantity of oxidizable material in a sample in all circumstances, e.g. in the presence of substances toxic to the oxidizing bacteria. In addition, measuring the *BOD* of a sample of water typically takes a minimum of five days. The chemical oxygen demand (*COD*) test is a quicker method, and measures the total oxygen demand. It is a measure of the total amount of oxygen required to stabilize all the waste. While the value of *COD* is never less than the value of *BOD*, it is the faster reacting *BOD* that impacts water quality. And this is what most people care about. However, determining a relationship between *BOD* and *COD* at any site can provide guideline values for *BOD* based on *COD* values.

Tables 13.8 and 13.9 provide some general ranges of pollutant concentrations in *CSO*s from urban catchments.

Water quality models for urban drainage are similar to, or simpler versions of, water quality models for other water bodies (as discussed in Chapter 12). Often a simple mixing and dilution model will be sufficient to predict the concentration of pollutants at any point in a sewage system. For example, such models may be sufficient for some toxic substances that are not broken down. Once the flows in the *CSO* enter the receiving water body, models discussed in Chapter 12 can be used to estimate their fate as they travel with the water in the receiving water body.

A factor that makes predicting the impacts of overflow discharges particularly difficult is the non-continuous nature of the discharges and their pollutant concentrations. In the first sanitary or foul flush, the fine sediments deposited in the pipes during dry periods are swept up and washed out of the system. Most existing models, termed constant concentration models, do not account for this phenomenon. Since many of the most significant pollution events occur when the river has low flows, and hence low dilution factors, the quantity of spill in the first flush may be very important in the overall pollution impact.

| constituent | raw | treated | overflow |
|---|---|---|---|
| suspended solids mg/l | 300 | 5 | 200 |
| *BOD* mg/l | 367 | 10 | 300 |
| *COD* mg/l | 470 | 15 | 350 |
| ammonia mg/l | 39 | 7 | 30 |

E020903r

**Table 13.8.** Typical quality of domestic sewage.

### 4.4.4. Bacteriological and Pathogenic Factors

The modelling of pathogenic microorganisms is particularly difficult since there are a very large number of pathogenic organisms, each usually with a unique testing procedure, many of which are expensive. Also many pathogens may present a significant risk to human health in very small numbers. Incubation periods of over twenty-four hours are not uncommon and there are as yet no automatic real-time monitoring techniques in commercial use.

The detection of pathogens relies heavily on indicator organisms that are present in feces in far higher numbers than the pathogenic organisms. Escherichia Coliform (E. coli) bacteria is the most common fecal indicator, and is used throughout the world to test water samples for fecal contamination.

Because of the problems in measuring micro-biological parameters involved in *CSO*s, most sophisticated methods of determining the quality of sewer water restrict themselves to more easily measured determinants such as *BOD*, *COD*, suspended solids, ammonia, nitrates and similar constituents.

### 4.4.5. Oil and Toxic Contaminants

Oils are typically discharged into sewers by people and industries, or are picked up in the runoff from roads and road accidents. Since oil floats on water surfaces and disperses rapidly into a thin layer, a small quantity of oil discharged into a water body can prevent reoxygenation at the surface and thus suffocate the organisms living there. The dispersal rate changes with oil viscosity, and

| constituent | highway runoff | residential area | commercial area | industrial area |
|---|---|---|---|---|
| suspended solids mg/l | 28 - 1178 | 112 - 1104 | 230 - 1894 | 34 - 374 |
| *BOD* mg/l | 12 - 32 | 7 - 56 | 5 - 17 | 8 - 12 |
| *COD* mg/l | 128 - 171 | 37 - 120 | 74 - 160 | 40 - 70 |
| ammonia mg/l | 0.02 - 2.1 | 0.3 - 3.3 | 0.03 - 5.1 | 0.2 - 1.2 |
| lead mg/l | 0.15 - 2.9 | 0.09 - 0.44 | 0.1 - 0.4 | 0.6 - 1.2 |

E020903s

**Table 13.9.** Pollutant concentrations (mg/l) in urban runoff.

the length of time the oil is a problem will partly depend on the surface area of the receiving water as well.

There are three main sources of toxic contaminants that may be discharged from *CSO*s:

- *Industrial effluents*. These could be anything from heavy metals to herbicides.
- *Surface washoff contaminants*. These may be contaminants washed off the surface in heavy rainstorms, and in agricultural and suburban residential areas will probably include pesticides and herbicides. In many cases these contaminants make a larger contribution to the pollutant load than the domestic sanitary flow.
- *Substances produced naturally in the sewer*. Various poisonous gases are produced in sewers. From the point of view of water quality, ammonia is almost certainly the most important, though nitrogen sulphide can also be significant.

#### 4.4.6. Suspended Solids

Discharges high in suspended solids pose a number of problems. They almost invariably exert an oxygen demand. If they remain in suspension they can prevent light from penetrating the water and thus inhibit photosynthesis. If deposited they become a reservoir of oxygen-demanding particles that can form an anaerobic layer on the bed, decreasing biodiversity. They can also degrade the bed for fish (such as salmon) spawning. If these suspended solids contain toxic substances, such as heavy metals, the problems can be more severe and complex.

## 5. Urban Water System Modelling

Optimization and simulation models are becoming increasingly available and are used to analyse a variety of design and operation problems involving urban water systems. Many are incorporated within graphic–user interfaces that facilitate the use of the models and the understanding and further analysis of their results.

### 5.1. Model Selection

A wide range of models is available for the simulation of hydrodynamics and water quality in urban systems. The selection of a particular model and the setup of a model schematization depends on the research question at hand, the behaviour of the system, the available time and budget, and future use of the model.

The research question and the behaviour of the water system determine the level of detail of the model schematization. The time scale of the dominating processes and the spatial distribution of the problem are key elements in the selection of a model, as is illustrated in Figure 13.18 and Table 13.10.

Figure 13.18 shows the time scales of the driving forces and their impact in urban water systems. It may be wise to consider the processes with largely different time scales separately, rather than joining them together in one model. For instance, the water quality of urban surface waters is affected by combined sewer overflows and by many diffusive sources of pollution. A combined sewer overflow lasts several hours and the impact of the discharge on the oxygen concentration in the surface water lasts for a couple of days. The accumulation of heavy metals and organic micro-pollutants in the sediment takes many years and the influx of the diffusive sources of pollution is more or less constant in time. The impact of combined sewer overflows on the oxygen concentration can be studied with a detailed, deterministic simulation model for the hydrodynamics and the water quality processes in the surface water system. A typical time step in such a model is minutes; a typical length segment is within the range from 10 to 100 metres. The accumulation of pollutants in the sediment can be modelled by means of a simple mass balance.

Another example is shown in Table 13.10. In this example the wastewater collection and treatment system in an urban area is modelled in three different ways. In the first approach, only the river is modelled. The discharge of effluent from a wastewater treatment plant is taken into account as a boundary condition. This is a useful approach for studying the impact of the discharge of effluent on water quality.

In the second approach, a detailed water and mass balance is made for an urban area. The main routes of water and pollution are considered. Generic measures, such as the disconnection of impervious areas from the sewage system, can be evaluated with this type of model schematization. In the third, most detailed, approach, a model schematization is made for the entire sewage system and, eventually, the wastewater treatment plant.

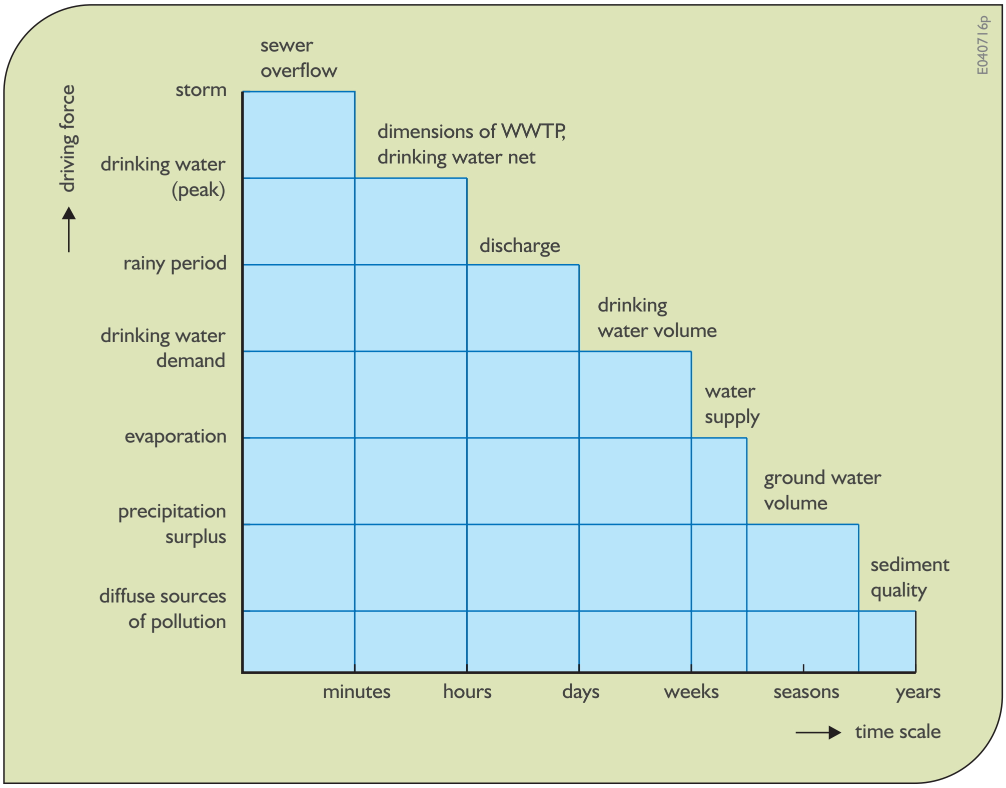



**Figure 13.18.** Time scales of driving forces and impacts in urban water systems.

| type of schematisation | implementation in code | examples |
|---|---|---|
| measured data of effluent quantity and quality | boundary conditions | end-of-pipe measures at the treatment plants |
| water balance and mass balance | rainfall runoff module, emissions module | analysis of main routes of rainfall discharge |
| detailed model schematisation | sewer flow module, water quality module | reduction of emissions via combined sewer overflows |

**Table 13.10.** Three methods of making a model schematization for an urban water system.

This type of model is needed when defining specific measures concerning storage in the sewage system, enhanced treatment capacity at the wastewater treatment plant and the design of stormwater retention ponds.

## 5.2. Optimization

The use of the storage, transport and treatment capacity of existing urban infrastructure can be optimized in many cases. Optimization of urban water systems aims at finding the technical, environmental and financial best solution, considering and balancing measures in the sewage system, the wastewater treatment plant and the surface water system at the same time. For instance, the optimum use of the storage capacity in a sewage system by means of real-time control of the pumps may eliminate the need for a more expensive increase in treatment capacity at the wastewater treatment plant. Storage

**Table 13.11.** Storage, transport and treatment capacity of sewers, wastewater treatment plants and surface waters.

| | sewer | WWTP | surface water |
|---|---|---|---|
| storage | moderate | none | limited - high |
| transport capacity | limited | limited | limited - high |
| treatment capacity | none | high | limited |

E040716r

of runoff from streets and roofs in surface water may be better for river water quality than transporting the runoff to the wastewater treatment plant and subsequent discharging it as effluent. Table 13.11 shows a matrix with the key variables of storage, transport capacity and treatment capacity in the sewage system, the wastewater treatment plant and the surface water.

Methods for finding optimal solutions are becoming increasingly effective in the design and planning of urban infrastructure. Yet they are challenged by the complexity and non-linearity of water distribution networks, especially urban ones.

Numerous calibration procedures for water distribution system models have been developed since the 1970s. Trial and error approaches (Rahal et al., 1980; Walski, 1983) were replaced with explicit type models (Boulos and Wood, 1990; Ormsbee and Wood, 1986). More recently, calibration problems have been formulated and solved as optimization problems. Most of the approaches used so far are either local or global search methods. Local search gradient methods have been used by Shamir (1974), Lansey and Basnet (1991), Datta and Sridharan (1994), Reddy et al. (1996), Pudar and Liggett (1992), and Liggett and Chen (1994) to solve various steady-state and transient model calibration problems (Datta and Sridharan, 1994; Savic and Walters, 1995; Greco and Del Guidice, 1999; Vitkovsky et al., 2000).

Evolutionary search algorithms, discussed in Chapter 6, are now commonly used for the design and calibration of various highly non-linear hydraulic models of urban systems. They are particularly suited for search in large and complex decision spaces, e.g. in water treatment, storage and distribution networks. They do not need complex mathematical matrix inversion methods and they permit easy incorporation of additional calibration parameters and constraints into the optimization process (Savic and Walters, 1995; Vitkovsky and Simpson, 1997; Tucciarelli et al., 1999; Vitkovsky et al., 2000).

In addition to calibration, these evolutionary search methods have been used extensively to find least-cost designs of water distribution systems (Simpson et al., 1994; Dandy et al., 1996; Savic and Walters, 1997). Other applications include the development of optimal replacement strategies for water mains (Dandy and Engelhardt, 2001), finding the least expensive locations of water quality monitoring stations (Al-Zahrani and Moied, 2001), minimizing the cost of operating water distribution systems (Simpson et al., 1999), and identifying the least-cost development sequence of new water sources (Dandy and Connarty, 1995).

These search methods are also finding a role in developing master or capital improvement plans for water authorities (Murphy et al., 1996; Savic et al., 2000). In this role they have shown their ability to identify low-cost solutions for highly complex water distribution systems subject to a number of loading conditions and a large number of constraints. Constraints on the system include maximum and minimum pressures, maximum velocities in pipes, tank refill conditions and maximum and minimum tank levels.

As part of any planning process, water authorities need to schedule the capital improvements to their system over a specified planning period. These capital improvements could include water treatment plant upgrades or new water sources as well as new, duplicate or replacement pipes, tanks, pumps and valves. This scheduling process requires estimates of how water demands are likely to grow over time in various parts of the system. The output of a scheduling exercise is a plan that identifies what facilities should be built, installed or replaced, to what capacity and when, over the planning horizon. This plan of how much to do and when to do it should be updated periodically long before the end of the planning horizon.

The application of optimization to master planning for complex urban water infrastructure presents a significant challenge. Using optimization methods to find the minimum-cost design of a system of several thousand pipes for a single demand at a single point in time is difficult enough on its own. The development of least-cost system designs over a number of time periods that experience multiple increasing demands can be much more challenging.

Consider, for example, developing a master plan for the next twenty years divided into four five-year construction periods. The obvious way to model this problem is to include the system design variables for each of the next four five-year periods, given the expected demands at those times. The objective function for this optimization model might be to minimize the present value of all construction, operation and maintenance costs. As mentioned previously, this is a very large problem that is probably unmanageable with the current state of technology for real water distribution systems.

Dandy et al. (2002) have developed and applied two alternative modelling approaches. One approach is to find the optimal solution for the system for only the final or 'target' year. The solution to this first optimization problem identifies those facilities that will need to be constructed sometime during the twenty-year planning period. A series of sub-problems are then optimized, one for each intermediate planning stage, to identify when each necessary facility should be built. For these sub-problems, the decisions are either to build or not to build to a predetermined capacity. If a component is to be built, its capacity has already been determined in the target year optimization.

For the second planning stage, all options selected in the first planning stage are locked in place and a choice is made from among the remaining options. Therefore, the search space is smaller for this case. A similar situation applies for the third planning stage.

An alternative approach is to solve the first optimization problem for just the first planning stage. All options and all sizes are available. The decisions chosen at this time are then fixed, and all options are considered in the next planning stage. These options include duplication of previously selected facilities. This pattern is repeated until the final 'target' year is reached.

Each method has its advantages and disadvantages. For the first, 'build-to-target' method, the optimum solution is found for the 'target year'. This is not necessarily the case for the 'build-up' method. On the other hand, the build-up method finds the optimal solution for the first planning stage, which the build-to-target method does not necessarily do. As the demands in the first planning stage are known more precisely than those for the 'target' year, this may be an advantage.

The build-up method allows small pipes to be placed at some locations in the first time planning stage, if warranted, and these can be duplicated at a later time; the build-to-target method does not. This allows greater flexibility, but may produce a solution that has a higher cost in present value terms.

The results obtained by these or any other optimization methods will depend on the assumed growth rate in demand, the durations of the planning intervals, the economic discount rate if present value of costs is being minimized, and the physical configuration of the system under consideration. Therefore, the use of both methods is recommended. Their outputs, together with engineering judgement, can be the basis for developing an adaptive master development plan. Remember, it is only the current construction period's solution that should be of interest. Prior to the end of that period, the planning exercise can be performed again with updated information to obtain a better estimate of what the next period's decisions should be.

## 5.3. Simulation

Dynamic simulation models are increasingly replacing steady-state models for analysing water quantity, pressure and water quality in distribution and collection networks. Dynamic models provide estimates of the time-variant behaviour of water flows and their contaminants in distribution networks, even arising from flow reversals. The use of long time-series analysis provides a continuous representation of the variability of flow, pressure and quality variables throughout the system. It also facilitates the understanding of transient operational conditions that may influence, for example, the way contaminants are transported within the network. Dynamic simulation also lends itself well to statistical analyses of exposure. This methodology is practical for researchers and practitioners using readily available hardware and software (Harding and Walski, 2002).

Models used to simulate a sequence of time periods must be capable of simulating systems that operate under highly variable conditions. Urban water systems are driven by water use and rainfall, which by their natures are stochastic. Changes in water use, control of responses and dispatch of sources, and random storms over different parts of the catchment, all can affect flow quantities and direction, and thus the spatial distribution of contaminants. Because different water sources often have different quality, changing water sources can cause changes in the quality of water within the system.

The simulation of water quantities and qualities in urban catchments serves three general purposes:

- *Planning/Design*. These studies define system configurations, size or locate facilities, or define long-term operating policies. They adopt a long-term perspective but, under current practices, use short, hypothetical scenarios based on representative operating conditions. In principle, the statistical distribution of system conditions should be an important consideration, but in practice variability is considered only by analyses intended to represent worst-case conditions.
- *Operations*. These short-term studies analyse scenarios that are expected to occur in the immediate future so as to inform immediate operational decisions. These are based on current system conditions and expected operating conditions. These analyses are often driven by regulations.
- *Forensics*. These studies are used to link the presence of contaminants to the risk or actual occurrence of disease. Depending on whether the objective is cast in terms of acute or chronic exposures, such studies may adopt short or long-term perspectives. Because there are often dose–response relationships and issues of latency in the etiology of disease, explicit consideration of the spatial distribution, timing, frequency, duration and level of contamination is important to these studies (Rodenbeck and Maslia, 1998; Aral, et. al., 1996; Webler and Brown, 1993).

# 6. Conclusions

Urban water systems must include not only the reservoirs, groundwater wells and aqueducts that are the sources of water supplies needed to meet the varied demands in an urban area, but also the water treatment plants, the water distribution systems that transport that water, together with the pressures required, to where the demands are located. Once used, the now wastewater needs to be collected and transported to where it can be treated and discharged back into the environment. Underlying all of this hydraulic infrastructure and plumbing is the urban stormwater drainage system.

Well-designed and operated urban water systems are critically important for maintaining public health as well as for controlling the quality of the waters into which urban runoff is discharged. In most urban areas in developed regions, government regulations require designers and operators of urban water systems to meet three sets of standards. Pressures must be adequate for fire protection, water quality must be adequate to protect public health, and urban drainage of waste and stormwaters must meet effluent and receiving water body quality standards. This requires monitoring as well as the use of various models for detecting leaks and predicting the impacts of alternative urban water treatment and distribution, collection system designs and operating, maintenance and repair policies.

Modelling the water and wastewater flows, pressure heads and quality in urban water conveyance, treatment, distribution and collection systems is a challenging exercise, not only because of its hydraulic complexity, but also because of the stochastic inputs to and demands on the system. This chapter has attempted to provide an overview of some of the basic considerations used by modellers who develop computer-based optimization and simulation models for design and/or operation of parts of such systems. These same considerations should be in the minds of those who use such models as well. Much more detail can be found in many of the references listed below.

## 7. References

AL-ZAHRANI, M.A. and MOIED, K. 2001. *Locating water quality monitoring stations in water distribution systems.* Proceedings of the World Water and Environmental Resources Congress. Orlando, Fla., ASCE, May.

ARAL, M.M.; MASLIA, M.L.; ULIRSCH, G.V. and REYES, J.J. 1996. Estimating exposure to volatile organic compounds from municipal water supply systems: use of a better computer model. *Archives of Environmental Health*, Vol. 51, No. 4, pp. 300–9.

BEDIENT, P.B. and HUBER, W.C. 1992. *Hydrology and floodplain analysis*, 2nd ed. Reading, Mass., Addison-Wesley.

BOULOS, P.F. and WOOD, D.J. 1990. Explicit calculation of pipe-network parameters. *Journal of Hydraulic Engineering,* ASCE, Vol. 116, No. 11, pp. 1329–44.

CHIN, D.A. 2000. *Water-resources engineering*. Upper Saddle River, N.J., Prentice Hall.

DANDY, G.C. and CONNARTY, M. 1995. Use of genetic algorithms for project sequencing. In: M.F. Domenica (ed.), *Integrated water resources planning for the twenty-first century*. Proceedings of the 22nd Annual Conference, Water Resources Planning and Management Division. ASCE, Cambridge, Mass., May, 540–3.

DANDY, G.C. and ENGELHARDT, M. 2001. The optimal scheduling of water main replacement using genetic algorithms. *Journal of Water Resources Planning and Management,* ASCE, Vol. 127, No. 4, July/August, pp. 214–23.

DANDY G.; KOLOKAS, L.; FREY, J.; GRANSBURY, J.; DUNCKER, A. and MURPHY, L. 2002. Optimal staging of capital works for large water distribution systems. In: D.F. Kibler (ed.), *Proceedings of the ASCE Environmental and Water Resources Planning Symposium*, *Roanoke, Va., May, 2002.* Reston, Va., ASCE Press (CD-rom).

DANDY, G.C.; SIMPSON, A.R. and MURPHY, L.J. 1996. An improved genetic algorithm for pipe network optimization. *Water Resources Research*, Vol. 32, No. 2, pp. 449–58.

DATTA, R.S.N. and SRIDHARAN, K. 1994. Parameter estimation in water-distribution systems by least squares. *Journal of Water Resources Planning and Management,* ASCE, Vol. 120, No. 4, pp. 405–22.

GRECO, M. and DEL GUIDICE, G. 1999. New approach to water distribution network calibration. *Journal of Hydraulic Engineering,* ASCE, Vol. 125, No. 8, pp. 849–54.

GUJER, W.; HENZE, M.; MINO, T. and VAN LOOSDRECHT, M.C.M. 1999. Activated sludge model no. 3. Wat. Sci. Tech., Vol. 39, No. 1, pp. 183–93.

HARDING, B.L. and WALSKI, T.M. 2002. *Long time-series simulation of water quality in distribution systems.* Proceedings of the ASCE Environmental and Water Resources Planning Symposium. Roanoke, Va., May .

HENZE, M.; GUJER, W.; MINO, T.; MATSUO, T.; WENTZEL, M.C.; VAN MARAIS, G.M. and VAN LOOSDRECHT, M.C.M. 1999. Activated sludge model no. 2D, ASM2D. *Water Science and Tech*nology, Vol. 39, No. 1, pp. 165–82.

HUBER, W.C. and DICKINSON, R.E. 1988. Storm water management model, version 4: user's manual. EPA/600/3-88/001a (NTIS PB88-236641/AS). Athens, Ga., US EPA, Also see http://www.epa.gov/ednnrmrl/swmm/ (accesssed 11 November 2004).

HVITVED-JACOBSEN, T.; VOLLERTSEN, J. and TANAKA, N. 1998. Activated sludge model no. 3. *Water Science and Technology*, Vol. 38, No. 10, pp. 257–64.

LANSEY, K.E. and BASNET, C. 1991. Parameter estimation for water distribution networks. *Journal of Water Resources Planning and Management,* ASCE, Vol. 117, No. 1, pp. 126–44.

LIGGETT, J.A. and CHEN, L.C. 1994. Inverse transient analysis in pipe networks. *Journal of Hydraulic Engineering,* ASCE, Vol. 120, No. 8, pp. 934–55.

MAYS, L.W. (ed.). 1999. *Hydraulic design handbook*. New York, McGraw-Hill.

MAYS, L.W. (ed.). 2000. *Water distribution systems handbook*. New York, McGraw-Hill.

MAYS, L.W. 2005. *Water resources engineering*. New York, Wiley.

MURPHY, L.J.; SIMPSON, A.R.; DANDY, G.C.; FREY, J.P. and FARRILL, T.W. 1996. Genetic algorithm optimization of the Fort Collins-Loveland Water Distribution

Systems. In: *Specialty conference on computers in the water industry*, AWWA, Chicago, Ill., April, pp. 181–5.

ORMSBEE, L.E. and WOOD, D.J. 1986. Explicit pipe network calibration. *Journal of Water Resources Planning and Management*, Vol. 112, No. 2, 166–82.

PRICE, R.K. 2002. *Urban drainage modeling: course notes*. Delft, UNESCO International Institute for Infrastructural, Hydraulic and Environmental Engineering.

PUDAR, R.S. and LIGGETT, J.A. 1992. Leaks in pipe networks. *Journal of Hydraulic Engineering,* ASCE, Vol. 118, No. 7, pp. 1031–46.

RAHAL, C.M.; STERLING, M.J.H. and COULBECK, B. 1980. Parameter tuning for simulation models of water distribution networks. *Proceedings of Institution of Civil Engineers*, Part 2, Vol. 69, pp. 751–62.

REDDY, P.V.N.; SRIDHARAN, K. and RAO, P.V. 1996. WLS method for parameter estimation in water distribution networks. *Journal of Water Resources Planning and Management,* ASCE, Vol. 122, No. 3, pp. 157–64.

RODENBECK, S.E. and MASLIA, M.L. 1998. Groundwater modeling and GIS to determine exposure to TCE at Tucson. *Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management*, ASCE, Vol. 2, No. 2, pp. 53–61.

ROESNER, L.A.; ALDRICH, J.A. and DICKINSON, R.E. 1988. *Storm Water Management Model User's Manual, Version 4: Addendum I, EXTRAN, EPA/600/3-88/001b (NTIS PB88236658/AS)*. Athens, Ga., Environmental Protection Agency.

SAVIC, D.A. and WALTERS, G.A. 1995. *Genetic algorithm techniques for calibrating network models, Report No. 95/12*. Exeter, UK, University of Exeter, Centre for Systems and Control Engineering.

SAVIC, D.A. and WALTERS, G.A. 1997. Genetic algorithms for least-cost design of water distribution networks. *Journal of Water Resources Planning and Management*, ASCE, Vol. 123, No. 2, pp. 67–77.

SAVIC, D.A.; WALTERS, G.A.; RANDALL SMITH, M. and ATKINSON, R.M. 2000. *Cost savings on large water distribution systems: design through GA optimisation*. Proceedings of the Joint Conference on Water Resources Engineering and Water Resources Planning and Management. Minneapolis, Minn., ASCE, July.

SHAMIR, U. 1974. Optimal design and operation of water distribution systems. *Water Resources Research*, Vol. 10, No. 1, 1974, pp. 27–36.

SIMPSON, A.R.; DANDY, G.C. and MURPHY, L.J. 1994. Genetic algorithms compared to other techniques for pipe optimization. *Journal of Water Resources Planning and Management,* ASCE, Vol. 120, No. 4, July/August, pp. 423–43.

SIMPSON, A.R.; SUTTON, D.C.; KEANE, D.S. and SHERIFF, S.J. 1999. Optimal control of pumping at a water filtration plant using genetic algorithms. In: D.A. Savic and G.A. Walters (eds.), *Water industry systems: modelling and optimization applications*, Baldock, UK, Research Studies Press. Vol. 2, pp. 407–15.

TUCCIARELLI, T.; CRIMINISI, A. and TERMINI, D. 1999. Leak analysis in pipeline systems by means of optimal valve regulation. *Journal of Hydraulic Engineering,* ASCE, Vol. 125, No. 3, 1999, pp. 277–85.

USDA (US Department of Agriculture). 1972. *National engineering handbook*. Washington, D.C., USGPO.

VITKOVSKY, J.P. and SIMPSON, A.R. 1997. *Calibration and leak detection in pipe networks using inverse transient analysis and genetic algorithms, report no. R 157*. Adelaide, University of Adelaide, Department of Civil and Environmental Engineering.

VITKOVSKY, J.P.; SIMPSON, A.R. and LAMBERT, M.F. 2000. Leak detection and calibration using transients and genetic algorithms. *Journal of Water Resources Planning and Management,* ASCE, Vol. 126, No. 4, pp. 262–5.

WALSKI, T.M. 1983. Technique for calibrating network models. *Journal of Water Resources and Planning Management,* ASCE, Vol. 109, No. 4, pp. 360–72.

WEBLER, T. and BROWN, 1993. I.S. Exposure to tetrachloroethylene via contaminated drinking water pipes in Massachusetts: a predictive model. *Archives of Environmental Health*, Vol. 48, No. 5, pp. 293–7.

## Additional References (Further Reading)

AXWORTHY, D.H. and KARNEY, B.W. 1996. Modeling low velocity/high dispersion flow in water distribution systems. *Journal of Water Resources Planning and Management,* ASCE, Vol. 122, No. 3, pp. 218–21.

CHAUDRY, M.H. and ISLAM, M.R. 1995. Water quality modeling in pipe networks. In: E. Cabrera and A.F. Vela (eds.), *Improving efficiency and reliability in water distribution systems*. Dordrecht, the Netherlands, Kluwer Academic.

CLARK, R.M.; GRAYMAN, W.M.; MALES, R.M. and COYLE, J.M. 1986. Predicting water quality in distribution systems. In: *Proceedings of the distribution system symposium*, Minneapolis, Minn. AWWA, September 7–10, pp. 155–80.

CLARK, R.M.; MALES, R. and STEVIE, R.G. 1984. *Water supply simulation model*, Vols. I, II and III. Cincinnati, Ohio, USEPA.

GRAYMAN, W.M.; CLARK, R.M. and MALES, R.M.1988. Modeling distribution system water quality: dynamic approach. *Journal of Water Resources Planning and Management,* ASCE, Vol. 114, No. 3, pp. 295–312.

ISLAM, M.R. and CHAUDRY, M.H. 1998. Modeling of constituent transport in unsteady flows in pipe networks. *Journal of Hydraulic Engineering,* ASCE, Vol. 124, No. 11, pp. 1115–24.

LINGIREDDY, S. and ORMSBEE, L.E. 1998. *Optimal network calibration model based on genetic algorithms*. Lexington, Ky., University of Kentucky.

MALES, R.N.; CLARK, R.M.; WEHRMAN, P.J. and GATES, W.E. 1985. Algorithms for mixing problems in water systems. *Journal of Hydraulics Engineering,* ASCE, Vol. 111, No. 2, pp. 206–19.

MEADOWS, M.E.; WALSKI, T.M.; BARNARD, T.E. and DURRANS, S.R. 2004. *Computer applications in hydraulic engineering*, 6th Edn. Waterbury, Conn.

MORLEY, M.S.; ATKINSON, R.M.; SAVIC, D.A. and WALTERS, G.A. 2001. GAnet: genetic algorithm platform for pipe network optimisation. *Advances in Engineering Software*, Vol. 32, No. 6, pp. 467–75.

ORMSBEE, L.E. 1989. Implicit network calibration. *Journal of Water Resources Planning and Management,* ASCE, Vol. 115, No. 2, pp. 243–57.

ROSSMAN, L.A.; BOULOS, P.F. and ALTMAN, T. 1993. Discrete volume-element method for network water quality models. *Journal of Water Resources Planning and Management,* ASCE, Vol. 119, No. 5, pp. 505–17.

ROSSMAN, L.A.; CLARK, R.M. and GRAYMAN, W.M. 1994. Modeling chlorine residuals in drinking-water distribution systems. *Journal of Environmental Engineering,* ASCE, Vol. 120, No. 4, pp. 803–20.

SHAMIR, U. and HOWARD, C.D.D. 1968. Water distribution systems analysis. *Journal of the Hydraulic Division,* ASCE, Vol. 94, No. 1, pp. 219–34.

TANG, K.; KARNEY, B.; PENDLEBURY, M. and ZHANG, F. 1999. Inverse transient calibration of water distribution systems using genetic algorithms. In: D.A. Savic and G.A. Walters (eds.), *Proceedings of water industry systems: modelling and optimisation applications*, Baldock, UK, Research Studies Press. Vol. 1, pp. 317–26.

TODINI, E. 1999. Using a Kalman filter approach for looped water distribution network calibration In: D.A. Savic and G.A. Walters (eds.), *Proceedings of water industry systems: modelling and optimisation applications*. Baldock, UK, Research Studies Press. Vol. 1, pp. 327–36.

USEPA. 1992. *CEAM systems development life cycle methodology (SDLCM) statement of policy, standards, and guidelines: Version 1.00*. Athens, Ga., USEPA.

VIESSMAN, W. Jr. and WELTY, C. 1985. *Water management technology and institutions*. New York, Harper and Row.

# 14. A Synopsis

1. Meeting the Challenge 461

2. The Systems Approach to Planning and Management 461
   2.1. Institutional Decision-Making 462
   2.2. The Water Resources System 464
   2.3. Planning and Management Modelling: A Review 465

3. Evaluating Modelling Success 466

4. Some Case Studies 467
   4.1. Development of a Water Resources Management Strategy for Trinidad and Tobago 468
   4.2. Transboundary Water Quality Management in the Danube Basin 470
   4.3. South Yunnan Lakes Integrated Environmental Master Planning Project 473
   4.4. River Basin Management and Institutional Support for Poland 475
   4.5. Stormwater Management in the Hague in the Netherlands 476

5. Summary 478

6. Reference 478

# 14 A Synopsis

This book introduced some approaches to modelling in support of water resources systems planning and management. These modelling tools can help identify and evaluate alternative plans and management policies in terms of their various physical, economic, ecological and environmental, and social impacts. Model outputs – information relevant for deciding what to do and when and where – are based on model inputs. Model inputs include data, judgements and assumptions. Models provide a way to identify just which of these data, judgements and assumptions are critical and which are not. This 'sensitivity' information can help focus the political decision-making debate on issues that are most important in the search of alternatives that best meet desired objectives and goals.

## 1. Meeting the Challenge

At the beginning of this twenty-first century, water developers and managers are being challenged to meet ever-increasing demands for more reliable, cheaper and cleaner water supplies for socio-economic development and, at the same time, ensure a quality environment and a resilient, sustainable and biodiverse ecosystem. Problems associated with periodic droughts, floods, pollution, aging infrastructure, threatened riverine and coastal environments and degraded ecosystems, conflicts over multiple purposes or uses, and increasingly issues related to safety and security, are common throughout much of the world. While there are many different ways of meeting these challenges, it seems obvious that water and related land development projects and management practices within any river basin should fit into an integrated resources management plan for the entire basin, including, in some cases, its estuarine and coastal zones.

Systems approaches encompassing multiple disciplines and multiple spatial and temporal scales are needed to help define and evaluate alternative plans and policies for managing the interdependent components of water resources systems, including their watersheds, as an integrated system. The condition of river basins in the next ten, twenty or fifty years will largely depend on how well various water-related economic, environmental and social interests within these basins have been satisfied in an integrated, equitable and sustainable way.

## 2. The Systems Approach to Planning and Management

Systems approaches, including the use of extensions of the modelling methods introduced in this book, can provide an organized framework for resources management and for estimating the important hydrological, geomorphic, ecological, social and economic impacts and trends over relevant scales of space and time. Within a systems framework, multiple purposes can be investigated, tradeoffs among competing objectives may be identified and evaluated, potential adverse impacts can be assessed, and the various costs and benefits, however measured, of a project may be estimated and examined. This can all be done within a context or process that incorporates the concerns and desires of all those with an interest or stake in the outcome.

Quantitative models can help inform interested stakeholders and those individuals or agencies responsible for recommending or making decisions or policy. The merits and advantages, as well as the limitations, of various quantitative methods for analysing various planning or management issues are generally recognized throughout the water resources community. The assumptions and uncertainties associated with any model-generated impact predictions should be understood and considered by those using these model predictions.

## 2.1. Institutional Decision-Making

Due to institutional, fiscal and political factors, the approach to decision-making in government water planning and management agencies is often piecemeal, fragmented and local-project oriented. The resulting decisions or actions are not always as effective or efficient as they could be. Decisions made today at the local level, without consideration of how the entire system works and how it will respond to such decisions, can lead to problems for tomorrow. History points to many examples.

The cost of restoring much of the unique, but degraded, Everglades ecosystem in south Florida in the United States is now estimated to exceed US$8 billion. Without considering the root causes of ecological stress on the Everglades, much of that $8 billion may be poorly spent. Land development and a rising sea level may continue to encroach on the natural system, whether restored or not. This reduces the ability of water managers to create the regimes of hydrological flows, water levels, water quality and hydro-periods needed to protect and enhance the various habitats within the Everglades (Figure 14.1).

The decision over the past decades to reroute river sediment at the mouth of the Mississippi River away from its delta has resulted in a loss of delta land (Figure 14.2). This delta supports a diverse ecosystem and an infrastructure used to offload crude oil from tanker ships and pump it in from offshore oil platforms. An estimated US$14 billion will be needed to protect and restore that eroding and subsiding delta. In addition, the excessive nutrient loads in runoff from farmlands throughout much of the Mississippi Basin have resulted in a hypoxic (dead) zone of over 7,000 square miles (some 19,000 km$^2$) along the Louisiana coast in the Gulf of Mexico. Eliminating that dead zone will be a massive political challenge in a basin where land owners do not appreciate being told how to manage their lands and their runoff.

Attempts to re-establish fish passage around dams (Figure 14.3) throughout the Columbia River and its tributaries have already cost an estimated US$3,000 per salmon based on the current salmon population in the river.

Success in solving these and similar challenges will depend on the extent to which they are viewed and managed as part of the entire river basin, estuary and coastal region, and on the extent of involvement of all affected stakeholders.

These are among the many water resources management issues that can be studied through the use of quantitative modelling methods. Clearly one of the biggest issues, even for river basins located entirely within single countries but especially for international river basins, is the institutional structure and capacity needed to take advantage of a more comprehensive systems approach to water resources planning and management. How can multiple government agencies, citizen or community groups, non-governmental organizations and private companies – each with interests, responsibilities and authorities to address a local or limited range of water management issues and needs – generate broadly supported integrated regional solutions? Where there exist active and well-staffed river basin commissions (such as those for the Delaware, Danube, Mekong and Rhine Rivers, to mention a few),

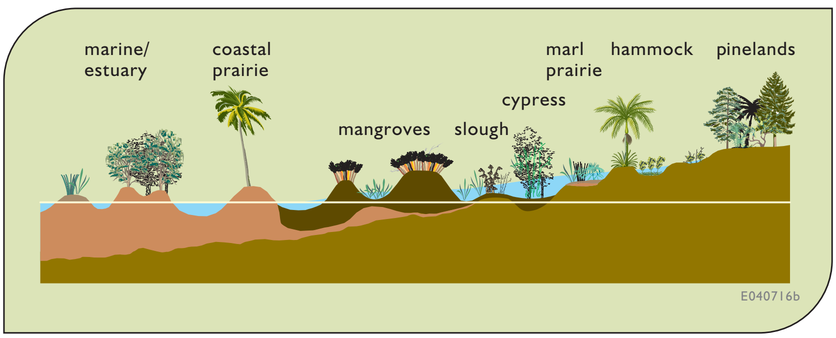



**Figure 14.1.** Ecosystem habitats in the Everglades in southern Florida (www.nps.gov).

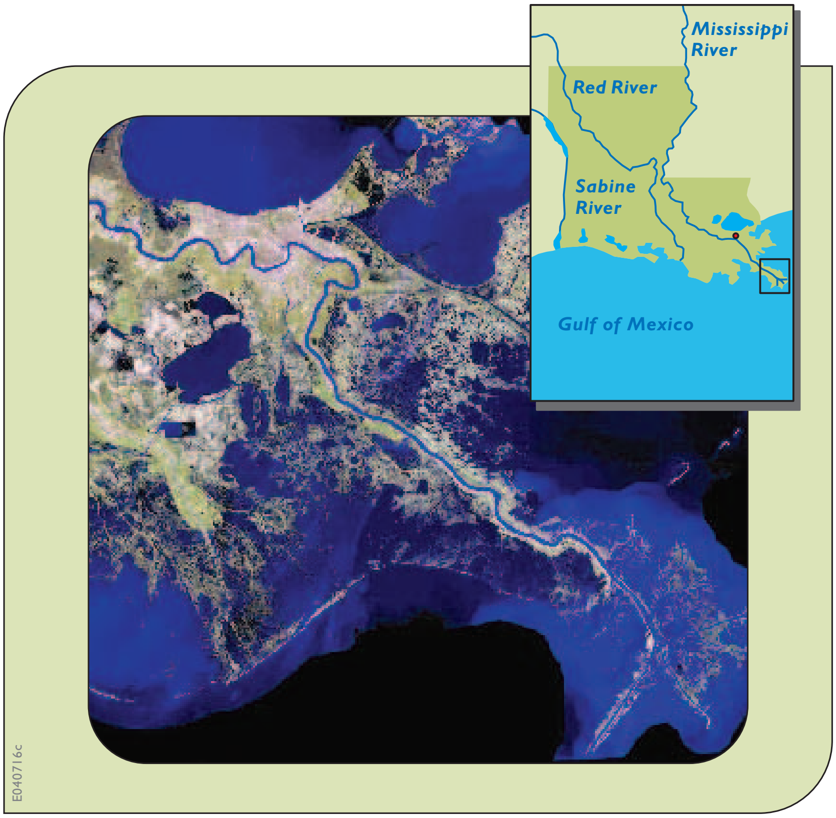



**Figure 14.2.** The Mississippi River Delta on the Louisiana coast in the Gulf of Mexico (NOAA Central Library, www.noaa.gov).

there is a greater potential for basin-wide coordination of multiple projects and management decisions. However, these commissions are not always able to fund the studies and analyses they would like to do in concert with other stakeholders in a participatory planning process that could result in strategies for guiding the integrated and sustainable development and management of their river basins (Chapter 1).

Water resources professionals and the informed public are increasingly seeking better ways of identifying and implementing solutions to water resources problems that can be done in less time and at a lower cost than traditional engineering projects. Working together in a participatory planning process, while less efficient in the speed of decision-making, is perhaps more effective. The more traditional top-down planning and management process is now becoming a multi-agency multi-stakeholder bottom-up planning and management process. Yet such a bottom-up process needs coordination to ensure unity of purpose, facilitate collaboration, deal with conflicts, and promote the inclusion of the full spectrum of public and private sector stakeholders. Both coordination and technical expertise and advice can be provided by government agencies if they have the ability and authority to do so. All too often, they do not.

Professionals in various disciplines must provide the information needed by those involved in the political planning and management process if they are to make


**Figure 14.3.** Obstacles to fish passage on the Columbia River in Canada and the United States.

informed decisions. This book has introduced some modelling approaches that are being used to provide some of this information. As discussed in Chapters 2 and 3, these tools can play an important role in helping stakeholders become better informed and reach a shared vision of how their water resources system functions and perhaps, if possible, even how they want it managed.

## 2.2. The Water Resources System

Understanding the natural processes as well as economic and social services or functions that rivers, lakes, reservoirs, wetlands, estuaries and coasts perform is critical to the successful and sustainable management of regional water resources systems. These natural processes involve numerous geomorphological, biological and chemical interactions that take place among the components of fluvial systems and their adjacent lands. Those who model and manage them should be aware of these processes and interactions.

The current hydrological, geomorphological, environmental and ecological states of streams, rivers, wetlands, estuaries and coasts are indicators of past and current management policies or practices. The condition of such components of a water resources system is an indicator of how well they can function. Natural systems are able to filter contaminants from runoff; store, absorb and gradually release floodwaters; serve as habitat for fish and wildlife; recharge groundwaters; provide for commercial transport of cargo; become sites for hydropower and provide recreational opportunities beneficial to humans. Degraded systems do not perform these functions so well, and thus can result in added costs of providing alternative ways of meeting these needs.

Today the importance of keeping natural aquatic systems alive and well, diverse and productive, in addition to meeting the needs of multiple economic and social interests, is much better appreciated and recognized than it was throughout much of our history up through the first half of the past century. During most of this time, water resources planners and managers were involved in 'conquering nature and taming its variability'. Today, natural system restoration and sustainability have become major objectives alongside the usual economic services that water and related land management can provide. Satisfying all these objectives to the extent possible requires an understanding of the basic hydrological, geological, environmental and ecological interactions that take place in natural aquatic systems.

Appendix A reviews some of the major natural processes and interactions that occur in watersheds, river basins, wetlands, estuaries and along coasts. It provides a background for those who may not be familiar with this subject and who wish to model such processes and their interactions for the purposes of planning and management.

## 2.3. Planning and Management Modelling: A Review

Water resources systems are characterized by multiple interdependent components that together produce multiple economic, environmental, ecological and social impacts. Planners and managers working toward improving the performance of, or the services provided by, these complex systems must identify and evaluate alternative plans and management or operating policies, comparing their predicted performance with desired goals or objectives. Constrained optimization and simulation modelling is the primary way we have of predicting system performance associated with any plan or operating policy. Chapters 3 and 4 introduced the development and use of some of the modelling methods commonly applied for identifying and analysing the performance of water resources systems. Chapter 5 showed how these models can be extended to include problems or issues better described qualitatively rather than quantitatively.

Models require data. Data sets contain information, often much more than can be learned from just looking at various graphical plots of those data. Statistical or data-based models help us abstract and gain new information and understanding from these data sets. They also serve as substitutes for more process-based models in applications where computational speed is critical or when more process-oriented models become too complex and/or too data demanding. Statistical data-based models range from the commonly used regression models to more recent ones based on evolutionary or biological processes. In Chapter 6 three such models were described. They are among a wide range of data mining tools designed to identify and abstract information from large sets of data.

Events that cannot be predicted precisely are often called random. Many if not most of the inputs to, and processes that occur in, water resources systems are to some extent random. Hence, so too are the outputs of predicted impacts, and even people's reactions to those outputs or impacts. To ignore this randomness or uncertainty when performing analyses in support of decisions is to ignore reality. Chapters 7, 8 and 9 reviewed some of the commonly used tools for dealing with uncertainty in water resources planning and management. Chapter 9 also addressed the topic of sensitivity analyses: ways of determining the sensitivity of model output to changes in the values of model inputs.

Water resources systems provide a variety of economic, environmental and ecological services. They also serve a variety of purposes (such as water supply, flood protection, hydropower production, navigation, recreation, and waste reduction and transport). Performance criteria provide measures of just how well a plan or management policy performs. There are a variety of criteria one can use to judge and compare system performance. Some of these performance criteria may be conflicting. The water that has access to one consumer is not available to another consumer, and vice versa. In these cases, tradeoffs exist among conflicting criteria and these tradeoffs must be considered when searching for the best compromise decisions. Chapter 10 presented ways of identifying and working with these tradeoffs in the political process of selecting the best decision.

Multipurpose river basin development projects typically involve the identification and use of both structural and non-structural measures that are designed to, for example, increase the reliability of municipal, industrial and agricultural water supplies when and where demanded, to protect against floods, to improve water quality, to provide for commercial navigation and recreation, and to produce hydropower, as appropriate for the particular river basin. Structural measures may include diversion canals, reservoirs, hydropower plants, levees, flood proofing, irrigation delivery and drainage systems, navigation locks, recreational facilities, groundwater wells, and water and wastewater distribution and collection systems and treatment plants. Non-structural measures may include land use controls and zoning, flood warning and evacuation measures, and economic and legal incentives that affect human behaviour with regard to water and watershed use. Planning the development and management of water resources systems involves identifying just what and when and where various structural or non-structural measures are needed, the extent to which they are needed, and their combined economic, environmental, ecological and social impacts. Chapters 11 and 12 introduced some modelling approaches for doing this. Chapter 11 focused on measures and models for water quantity management. Chapter 12 focused on measures and models applicable for water quality management.

Most people in most countries live in urban areas. These areas have their own special types of water resources management issues and problems. As populations continually move to urban areas to seek a higher standard of living and improved economic opportunities, and as cities merge to form megacities (with populations in excess of 10 million), the design and management of urban water supply systems becomes an increasingly important part of regional integrated water resources planning and management. Chapter 13 focused specifically on measures and models applied to the planning and management of water and wastewater in urban areas. Many of these tools are applicable in rural areas as well.

The management of water resource systems to serve various purposes and to meet various objectives requires actions with respect to land and water use. Regulating water quantity and quality often requires engineering infrastructure. The planning, design and operation of such infrastructure should be based on the best science available. Even the best science available is often not enough to give us the understanding we need to be sure we are making the right decisions. Furthermore, the goals and objectives our decisions are intended to meet can, and usually do, change over time. Given this uncertainty and changing political and social environment, it makes sense to monitor the impacts of our decisions and adapt – make changes – as appropriate.

All too often there is no follow-up monitoring carried out to judge the effectiveness of past decisions. Monitoring and adaptive management are increasingly being advocated when water management decisions are made on the basis of an uncertain science. But what should be monitored, and where, and how often and to what accuracy? These issues are addressed in Appendix B.

Considerable efforts have been (and continue to be) spent preparing for droughts or floods. Appendices C and D are specifically devoted to drought and flood management issues, where models can play an important role in evaluating the effectiveness of any measures taken to mitigate damage from these extreme hydrological events. Models and their associated computer software such as The Planning Kit of Delft Hydraulics for floodplain planning or the US Army Corps of Engineers 'shared vision' modelling approaches to drought management planning have proven successful at stakeholder, professional and various governmental levels in the Netherlands and in the US.

## 3. Evaluating Modelling Success

There are a number of ways one can judge the extent of success (or failure) in applying models and performing analyses in practice. Goeller (1988) suggested three measures as a basis for judging success:

1. How the analysis was performed and presented (analysis success).
2. How it was used or implemented in the planning and management processes (application success).
3. How the information derived from models and their application affected the system design or operation and the lives of those who use the system (outcome success).

It is often hard to judge the extent to which particular models, methods and styles of presentation are appropriate for the problem being addressed, the resources and time available for the study, and the institutional environment of the client. Review panels and publishing in peer-review journals are two ways of judging. No model or method is without its limitations. Two other obvious indications are the feelings that analysts have about their own work and, very importantly, the opinions the clients have about the analysts' work. Client satisfaction may not be an appropriate indicator if, for example, the clients are unhappy only because they are learning something they do not want to accept. Producing results primarily to reinforce a client's prior position or opinions might result in client satisfaction, but, most would agree, this is not an appropriate goal of modelling.

Application or implementation success implies that the methods and/or results developed in a study were seriously considered by those involved in the planning and management process. One should not, it seems to us, judge success or failure on the basis of whether or not any of the model results (the computer 'printouts') were directly implemented. What one hopes for is that the information and understanding resulting from model application helped define the important issues and identify possible solutions and their impacts. Did the modelling help influence the debate among stakeholders and decision-makers about what decisions to make or actions to take? The extent to which this occurs is the extent to which a modelling study will have achieved application or implementation success.

Outcome success is based on what happens to the problem situation once a decision largely influenced by the results of modelling has been made and implemented. The extent to which the information and understanding resulting from modelling helped solve the problems or resolve the issues, if it can be determined, is a measure of the extent of outcome success.

It is clear that success in terms of the second or third criteria will depend heavily on the success of the preceding one(s). Modelling applications may be judged successful in terms of the first two measures but, perhaps because of unpredicted events, the problems being addressed may have become worse rather than improved, or while those particular problems were eliminated, their elimination may have caused other severe problems. All of us can think of examples where this has happened. For example, any river restoration project involving the removal of engineering infrastructure is a clear indication of changing objectives or new knowledge. Who knows whether or not a broader systems study might have helped earlier planners, managers, and decision-makers foresee such consequences, but one cannot count on that. Hindsight is always clearer than foresight. Some of what takes place in the world is completely unpredictable. We can be surprised now and then. Given this, it is not clear whether we should hold modellers or analysts, or even planners or managers, completely responsible for any lack of 'outcome success' if unforeseen events that changed goals, or priorities or understanding did indeed take place.

Problem situations and criteria for judging the extent of success will change over time, of course. By the time one can evaluate the results, the system itself may have changed enough for the outcome to be quite different than what was predicted in the analysis. Monitoring the performance of any decision, whether or not based on a successfully analysed and implemented modelling effort, is often neglected. But monitoring is very important if changes in system design, management and operation are to be made to adapt to changing and unforeseen conditions. (Adaptive management is discussed in more detail in Appendix B.)

If the models, data, computer programs, documentation and know-how are successfully maintained, updated, and transferred to and used by the client institutions, there is a good chance that this methodology will be able to provide useful information relevant to the changes that are needed in system design, management or operation. Until relatively recently, the successful transfer of models and their supporting technology has involved a considerable commitment of time and money for both the analysts and the potential users of the tools and techniques. It has been a slow process. Developments in interactive computer-based decision support systems that provide a more easily understood human–model–data–computer interface have substantially facilitated this technology transfer process, particularly among model users. These technology developments have had a major impact on the state of the practice in using models in support of water resources planning and management activities.

## 4. Some Case Studies

What does one do with models used for analysing water resources planning and management issues? The following sections describe a few case studies or projects involving the application of modelling methods that are extensions of those described in this book. These projects were carried out by Delft Hydraulics in cooperation with other consulting firms and governmental and international institutions, and are typical of many such studies sponsored by and often carried out by international development banks and various UN and national foreign aid agencies. They serve to illustrate only a few of the wide variety of types of planning and management problems, issues, opportunities and challenges facing water resources planners and managers.

### 4.1. Development of a Water Resources Management Strategy for Trinidad and Tobago

The goal of this project was to develop a national water resources management strategy that would lead to sustainable management of the available water resources in Trinidad and Tobago. An important element in this project was the establishment of an effective and financially autonomous institutional setting that guaranteed the 'optimum management' of the islands' water resources sector. The project lasted just over one year. The planning horizons for the study were the years 2015 and 2025.

Trinidad and Tobago are two islands in the West Indies directly opposite the mouth of the Orinoco River (Figure 14.5). Although the average precipitation (2,000 mm/year) is quite high, its distribution in time (there is a wet and a dry season) and space (it is concentrated in the hills along the northern coast) does not match when and where it is needed. This mismatch between supply and demand creates a need for storage and distribution systems. The last decades have witnessed a deterioration of both availability and quality. The public water supply sector has failed to meet the demands for water during the wet as well as the dry season. Surface water quality poses a major constraint on the use of the resource. The lack of proper watershed management has resulted in increased peak flows and reduced discharge capacities of the rivers. Environmental problems have become increasingly apparent and will worsen if measures are not taken to reverse these trends. Without any major investments, the water resources system will limit the expected future growth in various demand sectors.

The project was divided into three distinctive phases (as described in the framework of analysis sections of Chapter 3 and further in Appendix E), each of which involved client participation. The Inception Phase defined the water resources system to be studied and identified the various problems and issues of concern. The Development Phase analysed the water-related problems and defined the scope of the study. This phase passed through a number of steps, from collection of data and information on all relevant subjects to a preliminary analysis of the water resources system. This analysis identified the various economic growth bottlenecks for present and future situations. Subsequently a large number of possible demand-oriented, supply-oriented, and environment-oriented measures were identified. These measures were grouped according to five development strategies that were considered as possible solutions for both present and future problems in the water resources situation. Each of these development strategies was examined with the aid of a river basin simulation model. A number of water supply, water quality, and environmental criteria were used to classify the degree of success of each strategy. These analyses included assessments of the necessary changes in

**Figure 14.4.** On the shores of Trinidad and Tobago, a hardship site for getting any work done!


**Figure 14.5.** Trinidad and Tobago are two islands located in the West Indies.

the institutional and legal framework that could lead to an improvement of the management of the water resources.

The water resources management study required an ample range of data, varying from discharge and rainfall data to per-capita water demand and expected transportation losses. The hydrological inputs to the simulation model were based on a statistical extension of the original historical records by applying time-series generation tools (as discussed in Chapter 7). The analysis included the assessment of the water availability (both surface water and groundwater) and the present and future (projected) water demand, both for public water supply (including industry) and irrigated agriculture.

The application of a simulation model allowed for the assessment of individual measures as well as groups of measures according to a certain development strategy. The impact of the various measures on the water flows and quality were displayed by means of the graphical tools in the simulation model. Figure 14.6 shows one such display.

An important aspect in the definition of a water resources strategy is the expected impact of any recommended measures on environmental and ecological conditions. Ongoing deterioration of the natural environment is evident from the major soil loss from hills due in part to the large-scale deforestation that has occurred in a number of regions.

On the basis of the information on water availability, water demand and other water-related aspects in Trinidad and Tobago, water resources management strategies were

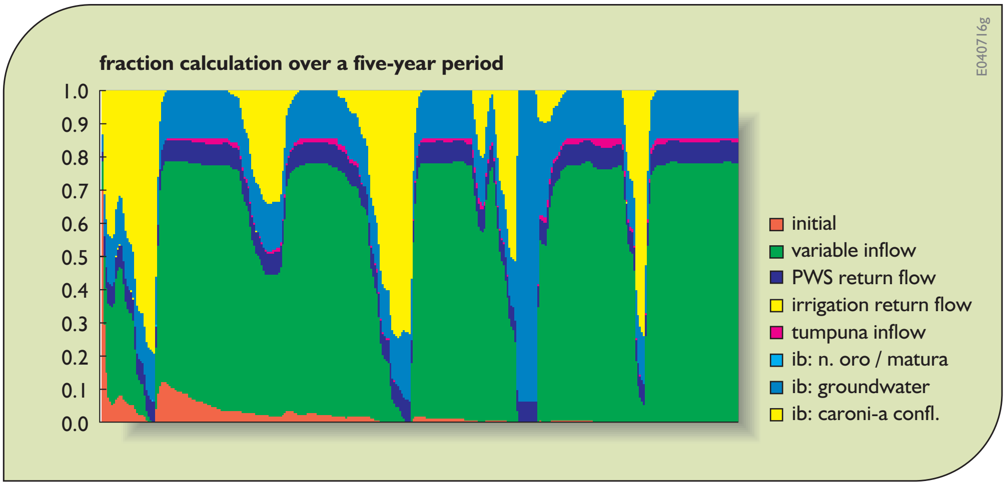



**Figure 14.6.** Composition of the water at the inflow to the Point Lisas Industrial Estate.

developed. They focused on different topics and goals. Each of the suggested strategies consisted of a number of related measures that in combination met pre-set criteria. These included supply reliability, environmental protection, government investment requirements and time required for implementation. Failure criteria were chosen such that both the time of failure and the extent of the failure would be taken into account. The final decision made by the government of Trinidad is illustrated in Figure 14.7.

Other aspects taken into account were changes in the existing legal framework, the impact on other agencies related to water resources, and the practical implementation of a new agency within the Ministry of Planning and Development. This agency should be financially independent, with income from licence fees covering the costs of the organization. A Planning Unit within that agency will have responsibility for the preparation and periodic revision of the water resources management plan, and will take over the planning activities that have been performed in the project.

## 4.2. Transboundary Water Quality Management in the Danube Basin

The Danube River Basin (as described in Chapter 1) covers some 817,000 square km in parts of eighteen countries in the heart of central Europe. The basin's population of 85-million is characterised by large socio-economic differences. The river flows from relatively rich western European states to relatively poorer former Soviet Union Republics. The river's larger tributaries include the Sava (Slovenia, Croatia, Bosnia-Herzegovina, Yugoslavia), the Tisa (Ukraine, Slovakia, Hungary, Romania, Yugoslavia), the Drava (Austria, Slovenia, Croatia, Hungary) and the Inn (Germany, Austria). The inhabitants of the basin use the Danube's water intensively (Figure 14.8). The basin includes many important natural areas, including the Danube delta, the second largest wetland area in Europe.

On 22 October 1998, the Danube River Protection Convention (DRPC) became the overall legal instrument for cooperation and transboundary water management in the Danube River Basin. The overall objective of the DRPC is to achieve and maintain the sustainable development and use of water resources in the Danube River Basin. The executing body of the convention is the International Commission for the Protection of the Danube River (ICPDR). A modelling project, lasting from 1995 to 2004, has assisted the ICPDR in establishing and implementing sustainable water quality management on a basinwide scale.

Events such as accidental spills, flooding and ice jams may degrade the downstream ecosystem and jeopardize the supplies of water. In such cases, there is a clear need for the collection and dissemination of early information about these events. To safeguard the aquatic environment and the use of water in the riparian states, the project developed and implemented an Accident Emergency Warning and Prevention System (AEWPS) for accidental pollutant discharges into the river Danube. The principal aim of the AEWPS is to communicate information about sudden changes in the water characteristics, such as


**Figure 14.7.** Water resources development strategy for Trinidad.

**Figure 14.8.** The 'Blue' Danube at Budapest, Hungary.

accidental spills or unpredictable changes in the water level, with special attention to transboundary impacts.

The focal locations of the AEWPS are the Principal International Alert Centres (PIACs), shown on Figure 14.9.

A Danube Basin Alarm Model (DBAM) was developed to support decision-making in relation to accidental spills with a probable transboundary impact. The January 2000 Baia Mare spill, Figure 14.10, presents a dramatic example.

The model provides forecasts of the travel time and the expected peak concentrations in the cloud of pollutants during its travel down the river. The DBAM was designed for use in operational conditions, to provide a fast and first-order assessment of the effects of a spill. It uses

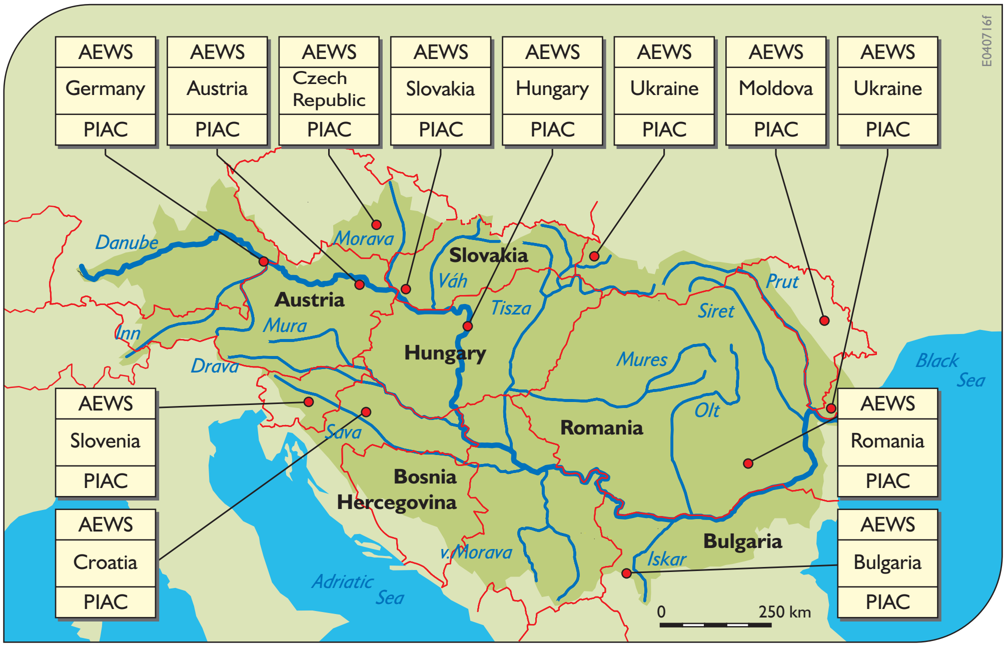



**Figure 14.9.** The Principle International Alert Centres (PIACs) of the Danube AEWPS.

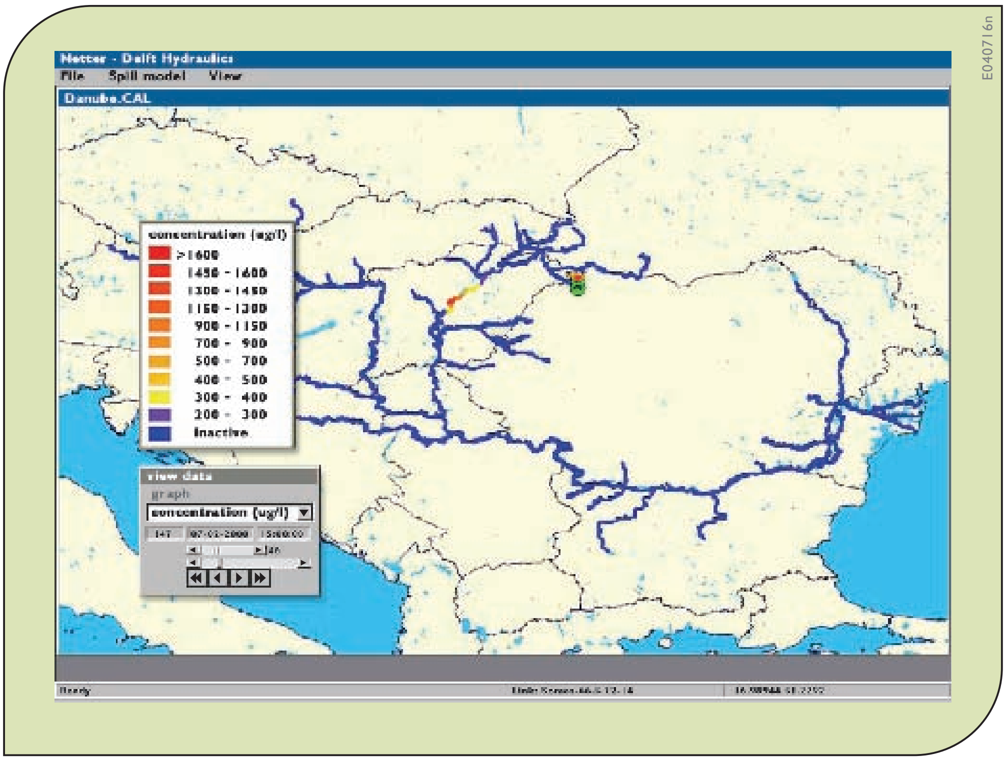



**Figure 14.10.** Application of the DBAM to the Baia Mare spill in January 2000.

readily available limited-input data. For reasons of computational speed and accuracy, the model uses analytical techniques to solve the governing mathematical advection–diffusion equation (discussed in Chapter 12). Experts in ten Danube countries are currently using the model. The water quality component of this model has been used to support a so-called Transboundary Diagnostic Analysis for the nutrients nitrogen and phosphorus, as well as to assess the effectiveness of the proposed pollution reduction program.

## 4.3. South Yunnan Lakes Integrated Environmental Master Planning Project

This project developed models and provided training aimed at protecting and rehabilitating four lakes: Fuxian, Xingyun, Qilu and Yilong. These four lakes are among the nine plateau lakes in Yunnan Province, China, (Figure 14.11). Figure 14.12 shows fishing boats on Fuxian Lake.

The South Yunnan lakes are threatened by water pollution and by decreased storage volumes, resulting in increased risks of flooding and drought. Water pollution is caused by wastewater discharges from industrial, urban

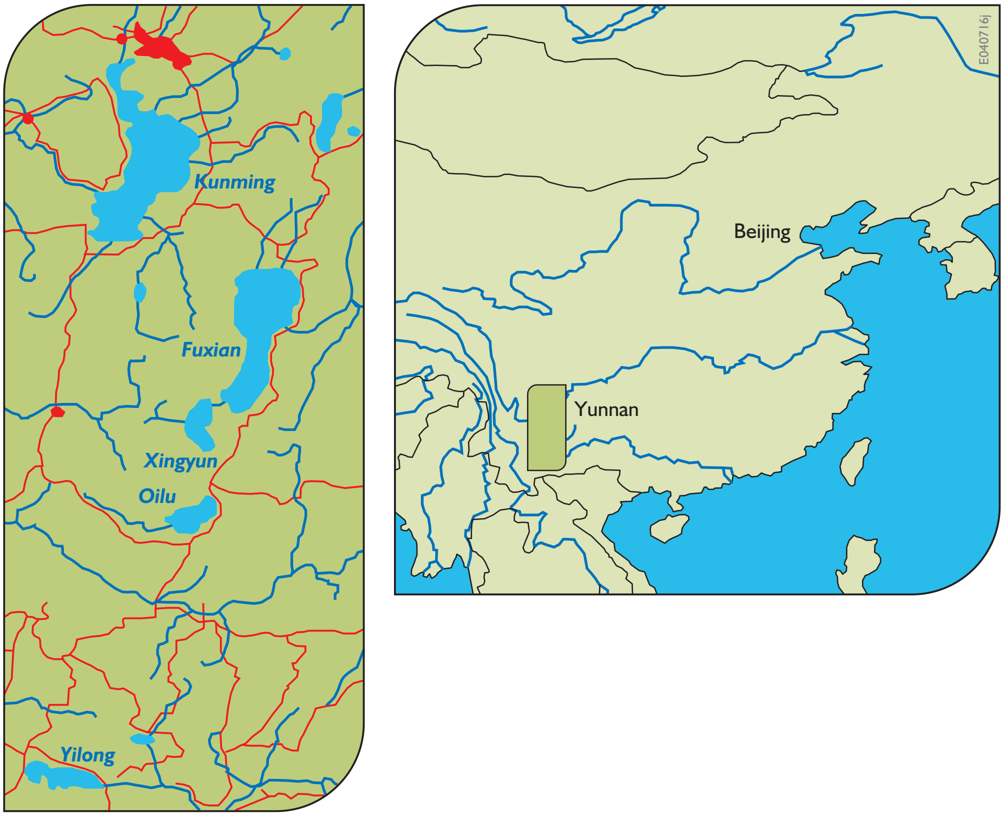



**Figure 14.11.** Lakes of project area in Yunnan Province, China.

**Figure 14.12.** Fishing boats on Fuxian Lake.

and agricultural sources. This leads to eutrophication of the lakes (high nutrient levels and high productivity of water plants and algae), as shown in Figures 14.13 and 14.14.

The decrease in storage capacity of the lakes is caused by the accumulation of organic material resulting from eutrophication, runoff and deposition of eroded soil material, and the reclamation of land for agricultural production. Soil erosion in the lake basins is a result of the collection of fuel wood, and overgrazing by livestock.

The South Yunnan Lakes Integrated Environmental Master Planning Project was aimed at reversing this trend, improving the quality of life in the lake basins, and demonstrating what could be done to restore the other lakes in the region. A decision support system (DSS) (Chapter 3) was developed to provide information on the effectiveness of various measures for accomplishing these project goals. These measures included:

- reduction of point and non-point pollution, to protect and improve the current water quality
- flood protection
- reduction of water shortages and the improvement of irrigation management and infrastructure (Figure 14.15)
- erosion prevention
- reduction of health impacts from pesticides, hazardous waste and bacteriological pollution
- institutional strengthening and awareness raising, leading to an integrated approach for the management of the lakes.

An evaluation process using multiple criteria analysis (Chapter 10) was used to select only the most effective measures. The first task of the DSS was to analyse the present water balance, waste loads and water quality. Second, it was used to estimate future conditions, including the lake quality, flood protection, reduction of water shortages, erosion prevention, biodiversity, public health, and macro economic development and poverty alleviation under different water management strategies or scenarios.

The criteria were weighted to reflect their relative importance to the development of the lake basins. The determination of the weighted factors, a political rather than technical issue, was determined for each lake during sessions held with the representatives of all involved at the Bureau and County levels (Figure 14.16).

The minimum package of measures needed to achieve the water quality targets was identified with the aid of the

**Figure 14.13.** Lotus flowers on Yilong Lake.

**Figure 14.14.** Children bathing among water hyacinths.

**Figure 14.15.** Water is pumped from Yilong Lake for irrigation.

WATER RESOURCES SYSTEMS PLANNING AND MANAGEMENT – ISBN 92-3-103998-9 – © UNESCO 2005

Figure 14.16. Workshop with local decision-makers about the weighting factors for the multiple criteria analysis.

DSS. This was followed by the selection of a package of additional measures to satisfy the other criteria, based on the multiple criteria analysis. Basic design reports were prepared for each of the selected measures, including the assessment of their technical, financial, economic and social feasibility and their environmental impact. Most importantly, institutional organizational issues were addressed and recommendations were made that would, if implemented, ensure the proper execution of the master plans and continued training, monitoring, planning and adaptation to changing goals and new knowledge.

## 4.4. River Basin Management and Institutional Support for Poland

A three-year project for the Polish Ministry of Environment, the agency responsible for water management in Poland, has involved river basin planning. The objective of this project was 'to support Poland with the implementation of the EU Water Framework Directive in Poland'. A pilot project in the Brda River catchment's area served to improve the skills of the Polish experts and other parties in the field of integrated water resources management.

The Brda River Basin is located in the northwest of Poland. The main management issue in the basin is the improvement of surface water quality to safeguard present and future drinking water abstractions from the river for the city of Bydgoszcz, shown in Figure 14.17.

The water quality problems in the basin are mainly due to agricultural and domestic wastewater discharges.

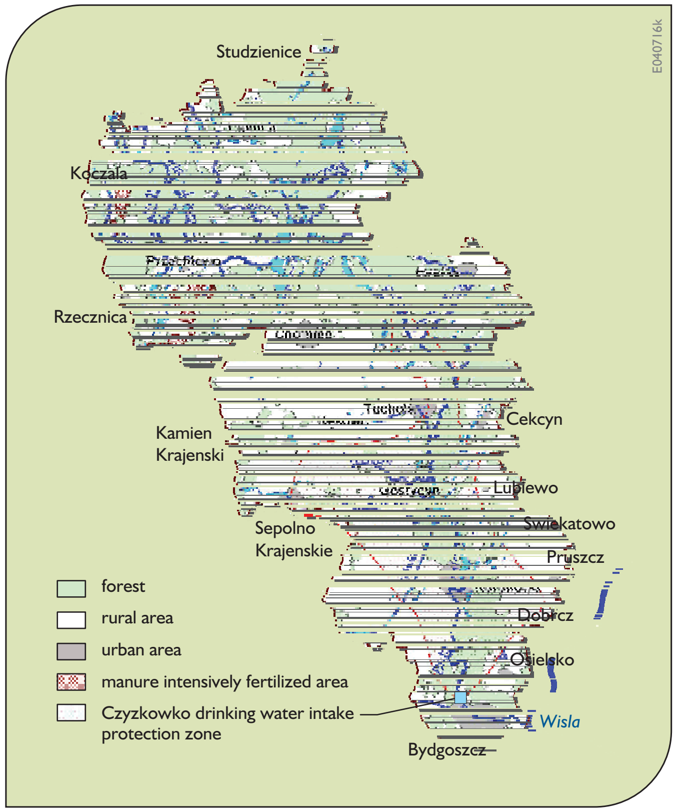



Figure 14.17. The Brda River Basin in Poland.

The process of preparing the entire plan for the Brda Basin, and implementing a training program, started at the beginning of 2001 and was completed in June 2003.

The development of an integrated water resources management plan for the Brda River (Figure 14.18) served to: first, strengthen the institutional organization and coordination; second, enhance public interest through an 'open planning process'; and third, improve the financial and operational management and the planning capacity of the ministry.

The Brda Catchment has a population of about 550,000, of which some 300,000 live in the area of Bydgoszcz, which is in the lower part of the Brda Basin. The upstream part of the basin is rural, with agricultural, forestry and some tourism activities, as shown in Figure 14.17. Forest and agriculture lands occupy about 85% of the basin area; 10% is urban, and 5% is covered by water bodies. Agriculture, farming and aquaculture are quite intensive, with many cattle (60,000), pigs (340,000), poultry (270,000) and sheep (10,000), as well as fish farms. This intensive agriculture causes water

**Figure 14.18.** Beeches by the Brda river (Marcin Kowalski).

quality problems that result from runoff of organic waste and nutrients. Industries are mainly found in and around Bydgoszcz. They include food production, (basic) metals and ceramics, and chemical industries (pharmaceutical, organic-chemical and the like).

Increasing agricultural and industrial activities are likely to result in increased animal waste loads. An increase of population will also put additional pressure on the water quality and the environment. The action plan is designed to protect the raw water source and safeguard the public water supply intake on Brda River for Bydgoszcz City. To estimate and evaluate the impacts of suggested projects and measures, computer simulations of the Brda Basin were executed. The results of these modelling studies were used by the Polish authorities to identify a final list of protective measures for the basin.

## 4.5. Stormwater Management in The Hague in the Netherlands

During periods of heavy rainfall when sewers are not capable of transporting all the stormwater to wastewater treatment plants, the excess water is discharged into the surface water. These sewer overflows degrade the water quality of urban channels and pools and flood the streets (Figure 14.19).

The objective of this study was to develop a set of measures with which both the emission standards and water quality standards for surface water could be met, at the lowest possible cost. The measures refer to both the sewage systems and the surface water.

A systems analysis was performed for the sewage and surface water networks of the Hague and its neighbouring municipalities. Discharges of sewage systems have to meet the Dutch emission standards and in addition must not degrade the quality of the receiving surface water. Attention was also given to preventing street flooding. The project scope included the municipalities of the Hague, Rijswijk, Leidschendam/Voorburg and Wateringen, and the water control board of Delfland. These municipalities share a large part of their sewage systems. The water control board of Delfland is responsible for both the water quantity and water quality in this region.

In this project, the sewage and surface water systems were modelled together as a single system. Sewage systems transport wastewater (and rainwater in the case of for combined sewage systems) to the wastewater treatment plant (WWTP). An overloaded sewage system results in overflows into surface water, which harm its

Figure 14.19. The results of excess stormwater on the otherwise picturesque canals and streets of urban areas in the Netherlands.

water quality. In the city of The Hague, surface water has a major role in boat transport and aesthetics (including tourism).

Within the region studied, the sewage system (Figure 14.20) serves 36 drainage areas and consists of about 1,360 km of main sewage pipes, 18 storage areas, 220 sewage overflows, and about 30,000 inspection sites. The surface water system consists of a large part of Delflands 'boezem' (storage system) and eleven polders. It includes about 835 km of canals, 46 pumps, 131 orifices, 594 culverts and 42 km$^2$ of drained surface area.

The sewage system is divided into eight areas which together cover some 23,000 manholes and 27,000 pipes. In the surface water model (Figure 14.21) only those channels are taken into account which were thought to be important for judging the effect of sewage overflows. Although many smaller channels were not modelled, the surface water model is one of the largest used for water quality evaluations, having over 5,500 water quality segments.

Models for the sewage and surface water systems were used to evaluate the reference situation. The model of the sewage system provided insight into situations where streets are flooded and sewage overflows occur. The effect of sewage overflows on water quality was simulated with the surface water model. The analysis of the reference

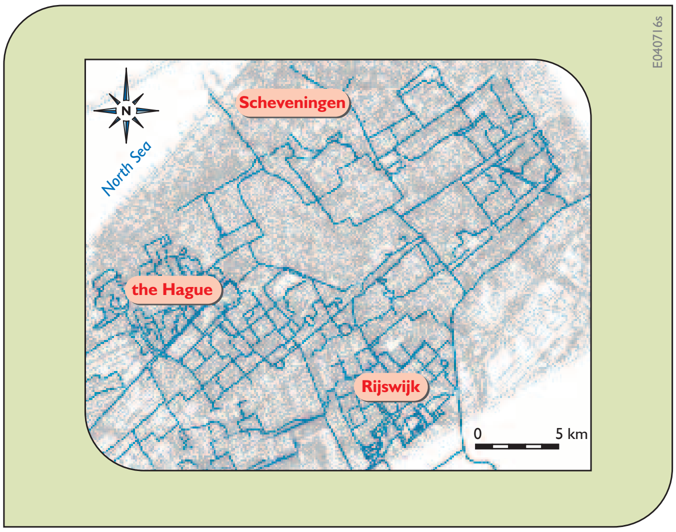



Figure 14.20. The main sewer system of the project area.

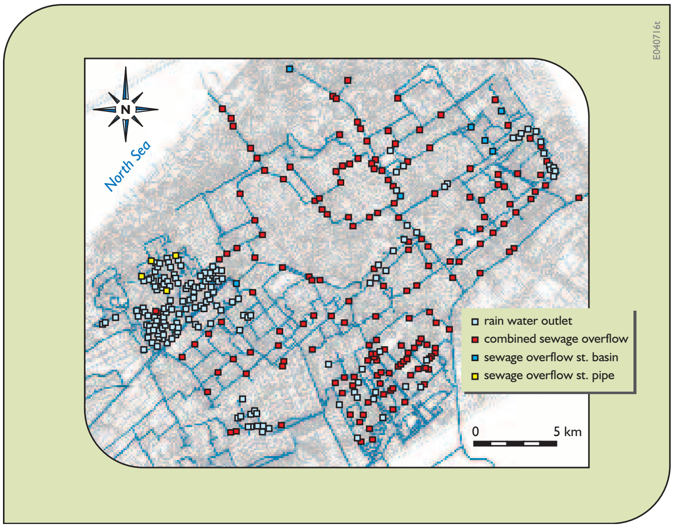



**Figure 14.21. The main surface water system of the project area.**

situation highlighted the major problem areas: water on streets, organic pollutant (BOD) loads, and water quality of receiving channels. The problems occurring in the receiving channels are caused by large residence times of the water and/or large BOD loads.

To analyse this problem, the following measures were considered:

- basic measures, aiming at compliance with emission standards
- integral measures, aiming at reduction of street flooding and compliance with the emission standards and the water quality requirements

Sewer system measures included:

- enlargement of the outflow capacity
- enlargement of the storage capacity
- construction of facilities at surface overflows, such as storage basins
- displacement of sewage overflows
- construction of separate or improved separate sewage systems
- transformation of combined sewer systems into improved separated sewage systems.

Surface water system measures included:

- enlargement of the surface water volume
- improvement of the circulation of the surface water for dilution and displacement of the BOD effluent load.

Combined system measures included:

- improving the discharge to existing storage basins
- addressing causes of sewage overflows with large BOD loads
- subdividing and separating drainage areas to reduce volumes of overflow at single discharge points.

Other measures involved the enlargement of pipes in the sewage system, displacement of sewage overflows, the enlargement of storage capacity in the sewage system, and the construction of flood control and sedimentation facilities at sewage overflows. Each of these measures aims at reducing the risk of flooding and BOD loads to the surface water.

The measures for surface water aim at reducing the effect of sewage overflows on the water quality. This may be done by improving the circulation of water near sewage overflows. Another measure prevents polluted water from flowing to a part of the surface water system that has a better water quality.

The models that were used in this project will be used in future evaluations of alternative measures to further improve the surface water quality in one of the largest cities of the Netherlands.

## 5. Summary

This concluding chapter has attempted to summarize and, using some case studies, illustrate the modelling methods introduced in this book for water resources systems planning and management. It is difficult today to find any water resources planning or management project which does not benefit from the use of modelling. Readers are encouraged to refer to the current journals and other sources of literature to learn more about this subject, and then enjoy the professional life of a water resources systems planner and manager. For us it has been fun, always challenging, and a profession that can serve and benefit society as well.

## 6. Reference

GOELLER, B.F., 1988. *A framework for evaluating success in systems analysis*. Report P-7454. Santa Monica, Calif., Rand Corporation.

## Appendix A: Natural System Processes and Interactions

1. Introduction 483

2. Rivers 483
   - 2.1. River Corridor 484
     - 2.1.1. Stream Channel Structure Equilibrium 485
     - 2.1.2. Lateral Structure of Stream or River Corridors 486
     - 2.1.3. Longitudinal Structure of Stream or River Corridors 487
   - 2.2 Drainage Patterns 488
     - 2.2.1. Sinuosity 489
     - 2.2.2. Pools and Riffles 489
   - 2.3. Vegetation in the Stream and River Corridors 489
   - 2.4. The River Continuum Concept 490
   - 2.5. Ecological Impacts of Flow 490
   - 2.6. Geomorphology 490
     - 2.6.1. Channel Classification 491
     - 2.6.2. Channel Sediment Transport and Deposition 491
     - 2.6.3. Channel Geometry 493
     - 2.6.4. Channel Cross Sections and Flow Velocities 494
     - 2.6.5. Channel Bed Forms 495
     - 2.6.6. Channel Planforms 495
     - 2.6.7. Anthropogenic Factors 496
   - 2.7. Water Quality 497
   - 2.8. Aquatic Vegetation and Fauna 498
   - 2.9. Ecological Connectivity and Width 500
   - 2.10. Dynamic Equilibrium 501
   - 2.11. Restoring Degraded Aquatic Systems 501

3. Lakes and Reservoirs 504
   - 3.1. Natural Lakes 504
   - 3.2. Constructed Reservoirs 505
   - 3.3. Physical Characteristics 505
     - 3.3.1. Shape and Morphometry 505
     - 3.3.2. Water Quality 506
     - 3.3.3. Downstream Characteristics 507
   - 3.4. Management of Lakes and Reservoirs 508
   - 3.5. Future Reservoir Development 510

4. Wetlands 510
   - 4.1. Characteristics of Wetlands 511
     - 4.1.1. Landscape Position 512
     - 4.1.2. Soil Saturation and Fibre Content 512
     - 4.1.3. Vegetation Density and Type 512
     - 4.1.4. Interaction with Groundwater 513
     - 4.1.5. Oxidation–Reduction 513
     - 4.1.6. Hydrological Flux and Life Support 513

4.2. Biogeochemical Cycling and Storage 513
4.2.1. Nitrogen (N) 514
4.2.2. Phosphorus (P) 514
4.2.3. Carbon (C) 514
4.2.4. Sulphur (S) 514
4.2.5. Suspended Solids 514
4.2.6. Metals 515
4.3. Wetland Ecology 515
4.4. Wetland Functions 515
4.4.1. Water Quality and Hydrology 515
4.4.2. Flood Protection 516
4.4.3. Shoreline Erosion 516
4.4.4. Fish and Wildlife Habitat 516
4.4.5. Natural Products 516
4.4.6. Recreation and Aesthetics 516

5. Estuaries 516
5.1. Types of Estuaries 517
5.2. Boundaries of an Estuary 518
5.3. Upstream Catchment Areas 519
5.4. Water Movement 519
5.4.1. Ebb and Flood Tides 519
5.4.2. Tidal Excursion 520
5.4.3. Tidal Prism 520
5.4.4. Tidal Pumping 520
5.4.5. Gravitational Circulation 520
5.4.6. Wind-Driven Currents 521
5.5. Mixing Processes 521
5.5.1. Advection and Dispersion 522
5.5.2. Mixing 522
5.6. Salinity Movement 523
5.6.1. Mixing of Salt- and Freshwaters 523
5.6.2. Salinity Regimes 523
5.6.3. Variations due to Freshwater Flow 523
5.7. Sediment Movement 524
5.7.1. Sources of Sediment 524
5.7.2. Factors Affecting Sediment Movement 524
5.7.3. Wind Effects 525
5.7.4. Ocean Waves and Entrance Effects 525
5.7.5. Movement of Muds 526
5.7.6. Estuarine Turbidity Maximum 527
5.7.7. Biological Effects 527
5.8. Surface Pollutant Movement 528
5.9. Estuarine Food Webs and Habitats 528
5.9.1. Habitat Zones 529

5.10. Estuarine Services 531
5.11. Estuary Protection 531
5.12. Estuarine Restoration 533
5.13. Estuarine Management 533
5.13.1. Engineering Infrastructure 534
5.13.2. Nutrient Overloading 534
5.13.3. Pathogens 534
5.13.4. Toxic Chemicals 534
5.13.5. Habitat Loss and Degradation 534
5.13.6. Introduced Species 535
5.13.7. Alteration of Natural Flow Regimes 535
5.13.8. Declines in Fish and Wildlife Populations 535

6. Coasts 535
6.1. Coastal Zone Features and Processes 535
6.1.1. Water Waves 536
6.1.2. Tides and Water Levels 538
6.1.3. Coastal Sediment Transport 538
6.1.4. Barrier Islands 538
6.1.5. Tidal Deltas and Inlets 538
6.1.6. Beaches 538
6.1.7. Dunes 539
6.1.8. Longshore Currents 540
6.2. Coasts Under Stress 540
6.3. Management Issues 540
6.3.1. Beaches or Buildings 542
6.3.2. Groundwater 542
6.3.3. Sea Level Rise 542
6.3.4. Subsidence 543
6.3.5. Wastewater 544
6.3.6. Other Pollutants 544
6.3.7. Mining of Beach Materials 545
6.4. Management Measures 545
6.4.1. ‘Conforming Use’ 546
6.4.2. Structures 546
6.4.3. Artificial Beach Nourishment 547

7. Conclusion 548

8. References 549

# Appendix A: Natural System Processes and Interactions

Understanding the natural processes as well as the economic and social services or functions that rivers, lakes, reservoirs, wetlands, estuaries and coasts fulfill is critical to the successful and sustainable management of these hydrological systems. These natural processes involve numerous physical and biological interactions that take place among the components of fluvial systems and their adjacent lands. This appendix briefly views some of the important natural processes and interactions that occur in watersheds, river basins, estuaries and coasts. Those who manage them should be aware of these processes and interactions as they build and use their models to analyse, plan and evaluate alternative management policies and practices.

## 1. Introduction

The hydrological, geomorphological, environmental and ecological state of streams and rivers, and of their downsteam estuaries and coasts, are indicators of past and current management policies or practices. The condition of a stream, river, estuary or coast is an indicator of how well it, as a system, can function. Natural systems can filter contaminants from runoff; store, absorb and gradually release floodwaters; serve as habitat for fish and wildlife; recharge groundwater; provide for commercial transport of cargo; become sites for hydropower; and provide recreational opportunities beneficial to humans. Degraded systems do not perform these functions as well as non-degraded systems.

Today the importance of keeping natural aquatic systems alive and well, diverse and productive, in addition to meeting the needs of multiple economic and social interests, is much better appreciated and recognized than it was when water resources planners and managers were involved in 'conquering nature and taming its variabilities'. Natural system restoration and sustainability have become major management objectives, along with the maintenance of the usual economic services that water and related land management can provide. Satisfying all these objectives as far as possible requires an understanding of the basic hydrological, geological, environmental and ecological interactions that take place in natural aquatic systems.

This appendix is divided into five main sections. The next section will focus on rivers. It is followed by sections describing lakes and reservoirs, wetlands, estuaries, and finally coasts. Multiple books have been written on each of these water resources system components. What is presented in this appendix is thus only an outline of the main features of these water bodies and how they can be managed. Its purpose is to introduce some vocabulary and serve as a primer for those not familiar with this background information.

## 2. Rivers

Rivers are driven by hydrological, fluvial, geomorphologic, water quality and ecological processes that occur over a range of temporal and spatial scales. The plant and animal communities in rivers and their floodplains are dependent upon change: changing flows, moving sediments and

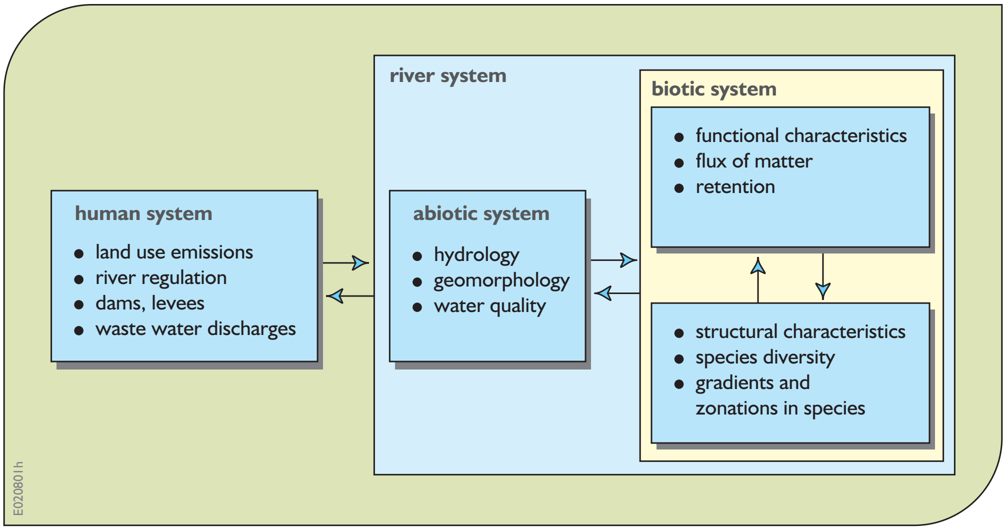



**Figure A.1.** Cause and effect chain of factors influencing a river system.

shifting channels. They depend on inputs of organic matter from vegetation in the riparian zone. They depend on the exchange of nutrients, minerals, organic matter and organisms between the river and its floodplain. This is provided by variable flows and sediment transport. All of these factors influence the structural character of the stream or river and its aquatic and terrestrial ecosystems – their species distribution, diversity and abundance.

Assessments of the effect of human activities on river systems require indicators relating cause to effect. A cause–effect chain is illustrated in Figure A.1. Insight into connections between processes and structures and their temporal and spatial scales leads to a more integrated interdisciplinary approach to river system monitoring and management.

## 2.1. River Corridor

A river corridor can be viewed as a hierarchical series of river segments, from upstream headwater streams to large downstream rivers, as illustrated in Figure A.2. River corridors include the river channels and the river margins (the water–land interfaces), and both are influenced by surface water–groundwater interactions. These environments are characterized by hydrological, geomorphological, environmental and ecological interactions. River margin interactions influence surrounding terrestrial landscapes.

Features that influence the structure and functioning of river systems occur at various spatial scales. A stream or river, for example, has an input–output relationship with the next higher scale, the stream or river corridor. This corridor scale, in turn, interacts with the landscape scale, and so on up the hierarchy. Similarly, because each larger-scale system contains the smaller scale ones, the structure and functions of the smaller systems affect the structure and functions of the larger.

Investigating relationships between structure and scale is a key first step for planning and designing stream or river system management plans. Landscape ecologists use four basic terms to define spatial structure at a particular scale.

These spatial landscape, component types in river basins, illustrated in Figure A.3 are:

- *Matrix*. The dominant and interconnected land cover in the basin.
- *Patch*. A different type of land cover found on smaller areas within the matrix.
- *Corridor*. A land cover type that links other patches in the matrix. Typically, a corridor is elongated in shape, such as a stream or river.
- *Mosaic*. A collection of isolated patches.

The 'watershed scale' that includes the stream corridor is a common scale of management, since many functions of the stream corridor are closely tied to drainage patterns. While the watershed scale is often the focus of river restoration and water resources management, especially for non-point pollutant discharge management, the other spatial scales should also be considered when developing a stream or river system management policy or plan. The

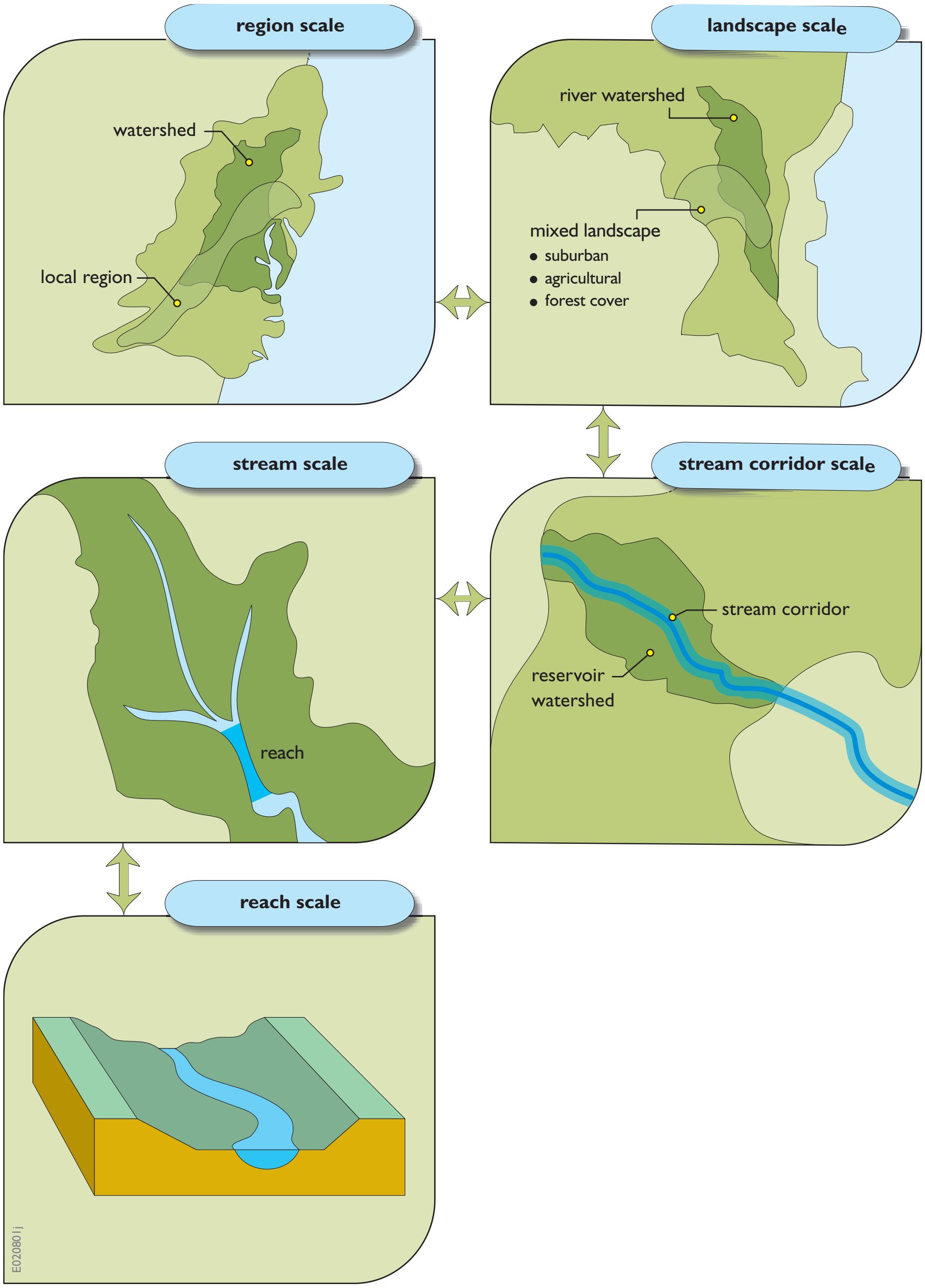



**Figure A.2.** River basin components viewed at multiple spatial scales, from the large regional scale to the local stream segment scale.

exclusive use of watersheds for the large-scale management of stream corridors, however, ignores the materials, energy and organisms that move across and through landscapes independent of water drainage. A more complete large-scale perspective of the stream and river system management is achieved when watershed hydrology is combined with landscape ecology and when actions in 'problem sheds' rather than only in drainage basins are being considered.

### 2.1.1. Stream Channel Structure Equilibrium

Nearly all channels are formed, maintained and altered by flows and sediment loads. Channel equilibrium involves the relation among four basic factors: sediment discharge, $Q_s$; sediment particle size, $D$; streamflow, $Q_w$; and stream slope, $S$. Lane (1955), using median particle size, $D_{50}$, expressed this relationship qualitatively as:

$$Q_s \cdot D_{50} \propto Q_w \cdot S \tag{A.1}$$

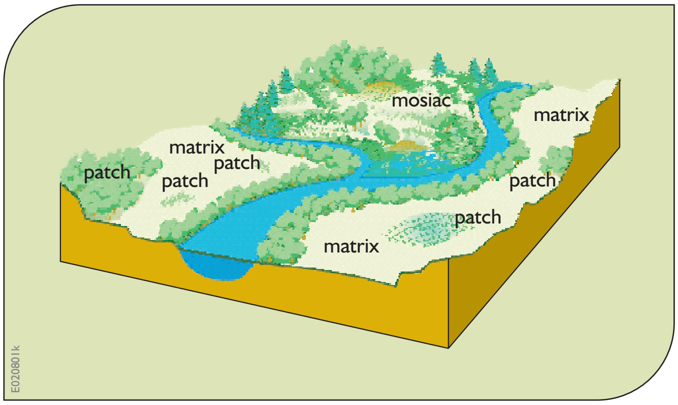



**Figure A.3.** River basin landscapes made up of matrix, patch, corridor and mosaic components at various scales.

This relationship states that a measure of sediment load (sediment discharge $Q_s$ times median particle size $D_{50}$) is proportional to a measure of the streamflow power (streamflow $Q_w$ times slope $S$). Channel equilibrium occurs when the streamflow power is constant over the length of the stream. If this occurs, no net changes in the channel shape will occur. If a change occurs in either the left or right-hand side of Equation A.1, the balance and hence equilibrium will be temporarily lost. If one variable changes, one or more of the other variables must change appropriately if equilibrium is to be maintained. Reaching equilibrium typically involves erosion and/or deposition.

Assuming increasing flows from runoff in the downstream direction, the channel slope has to be decreasing in the downstream direction. If the slope is too steep, sediment is deposited to reduce that steepness. This is why stream channels that experience increasing downstream flows have decreasing slopes in the downstream direction.

If streams in equilibrium have constant streamflow power, $Q_wS$, over distance, from Equation A.1, the sediment load, $Q_sD_{50}$, must also be constant. Hence, if sediment deposition is occurring in the downstream direction to decrease stream slopes, the median particle size, $D_{50}$, will be decreasing and the sediment discharge, $Q_s$, along with streamflow, $Q_w$, will be increasing. This is typically observed in channels with increasing downstream streamflows.

A stream seeking a new equilibrium tends to erode more sediment and larger particle sizes. Alluvial streams that are free to adjust to changes in these four variables generally do so and re-establish new equilibrium conditions. Non-alluvial streams such as those flowing over bedrock or in artificial, concrete channels are unable to maintain this equilibrium relationship because of their inability to pick up additional sediment.

The stream balance expressed in Equation A.1 can be used to make qualitative predictions about the impacts of changes in runoff or sediment loads from a watershed. Quantitative predictions, however, require the use of more complex simulation or physical models.

### 2.1.2. Lateral Structure of Stream or River Corridors

Stream and river valleys are created over time by the stream or river depositing sediment as it moves back and forth across the valley floor. These processes of lateral migration and sediment deposition, usually occurring during flood flows, continually modify the floodplain. Through time, as the channel migrates, it will maintain the same average size and shape as long as the channel stays in equilibrium.

One can distinguish two types of floodplains. The hydrological floodplain is the land adjacent to the baseflow channel residing below bank-full elevation. It is inundated about two years out of three. Not every stream corridor has a hydrological floodplain. The topographic floodplain is the land adjacent to the channel, including the hydrological floodplain, that is flooded by a flood peak of a given frequency (for example, the 100-year flood – the flood that is equalled or exceeded once every 100 years on average – defines the 100-year floodplain). Higher flood-peak flow return periods define wider topographic floodplains. These two types of floodplains are shown in Figure A.4.

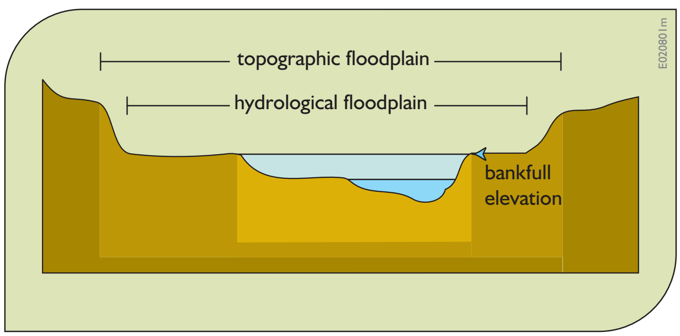



**Figure A.4.** Two types of floodplains, the hydrological and topographic.

Floodplains provide temporary storage space for floodwaters and sediment. This lengthens the lag time of a flood. This lag time is the time between the middle of the rainfall event and the runoff peak. If a stream's capacity for moving water and sediment is diminished, or if the sediment loads become too great for the stream to transport, the valley floor will begin to fill.

Topographic features on the floodplain, as illustrated in Figure A.5, include:

- *Meander scroll*. A sediment formation marking former channel locations.
- *Chute*. A new channel formed across the base of a meander. As it grows in size, it carries more of the flow.
- *Oxbow*. A severed meander after a chute is formed.
- *Clay plug*. A soil deposit at the intersection of an oxbow and the new main channel.
- *Oxbow lake*. A water body created after clay plugs separate the oxbow from the main channel.
- *Natural levees*. Formations built up along the bank of some streams that flood. As sediment-laden water spills over the bank, the sudden loss of depth and velocity causes coarser-sized sediment to drop out of suspension and collect along the edge of the stream.
- *Splays*. Delta-shaped deposits of coarser sediments that occur when a natural levee is breached. Natural levees and splays can prevent floodwaters from returning to the channel when floodwaters recede.
- *Backswamps*. A term used to describe floodplain wetlands formed by natural levees.

These different features provide a variety of habitats for plants and animals.

### 2.1.3. Longitudinal Structure of Stream or River Corridors

The processes that determine the characteristic lateral structure of a stream corridor also influence its longitudinal structure. For streams and rivers whose flows increase with distance downstream, channel width and depth also increase downstream due to increasing drainage area and discharge. Even among different types of streams, a common sequence of structural changes, as shown in Figure A.6, is observable from headwaters to mouth.

The longitudinal profile of many streams can be divided into three zones. The changes in the three zones are characterized in Figure A.6. Zone 1, the headwater zone, has the steepest slopes. In this zone sediments erode from slopes of the watershed and move downstream. The rivers in hilly regions are characterized by the swiftness of the flow in restricted and/or steep channels, the occurrence of landslides and the formation of rapids along their courses. The control of rivers in the upper reaches is known as 'torrent control'. Zone 2, the transfer zone, receives some of these sediments and hence is usually characterized by wider floodplains and more meandering channel patterns. The flatter slopes in zone 3 receive most of the coarser sediments. 'River training' methods are often adopted for managing alluvial rivers in this most downstream zone.

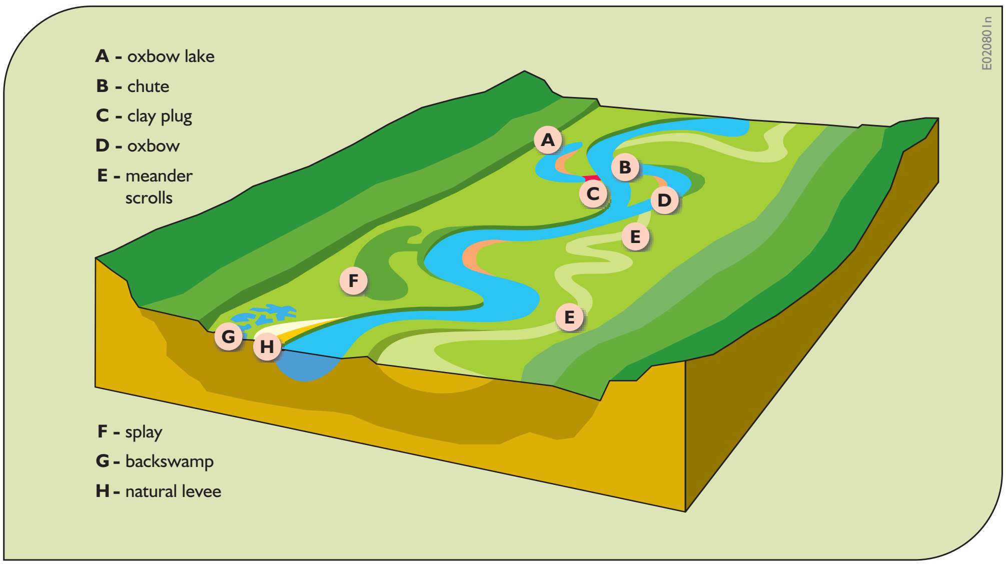



**Figure A.5.** Topographic features of a meandering stream on a floodplain.

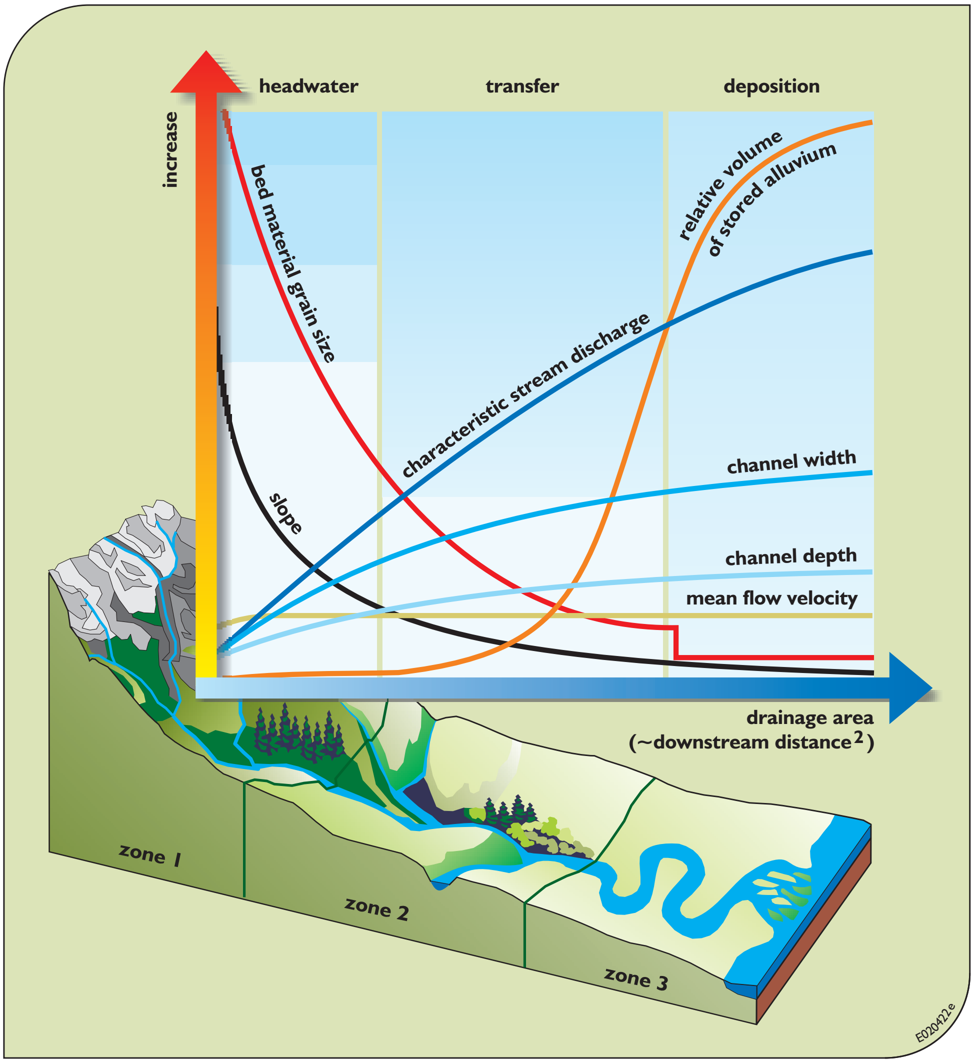



**Figure A.6.** Typical changes in the stream channel characteristics along its length.

Though Figure A.6 displays headwaters as mountain streams, these general patterns and changes apply to watersheds with relatively small topographic relief from the headwaters to the mouth. Erosion and deposition occur in all zones, but the zone concept focuses on the most dominant process.

## 2.2. Drainage Patterns

One distinctive aspect of a watershed or river basin when observed from above (bird's-eye view) is its drainage pattern. Drainage patterns are primarily controlled by topography and geologic structure. Figure A.7 shows a method of classifying, or ordering, the hierarchy of natural channels within a watershed or basin. This is a modified method based on the one proposed by Horton (1945).

The uppermost channels in a drainage network (headwater channels with no upstream tributaries) are designated as first-order streams down to their first confluence. A second-order stream is formed below the confluence of two first-order channels. Third-order streams are created when two second-order channels join, and so on. The intersection of a channel with another channel of lower order does not raise the order of the stream below the intersection, e.g. a fourth-order stream intersecting with, a second-order stream is still a fourth-order stream below the intersection.

Within a given drainage basin, stream order correlates well with other basin parameters, such as drainage area or channel length. Consequently, knowing what order a stream is can provide clues to other characteristics, such as its longitudinal zone and its relative channel size and depth.

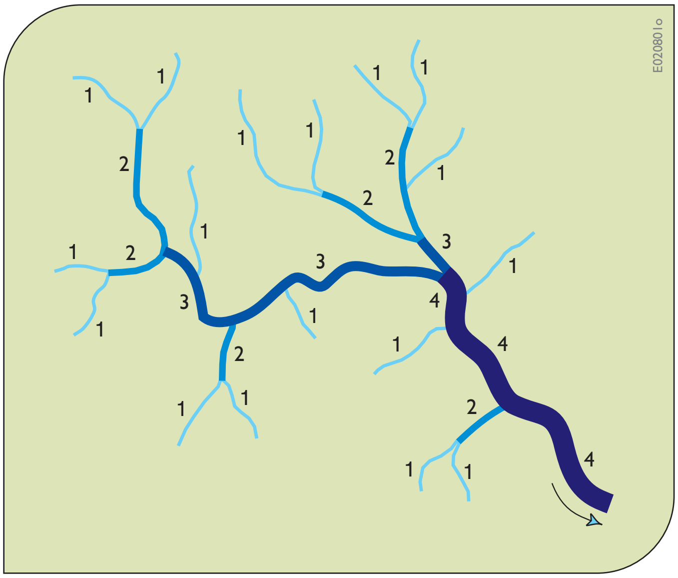



**Figure A.7.** Stream ordering in a drainage network showing first-order streams down to fourth-order streams.

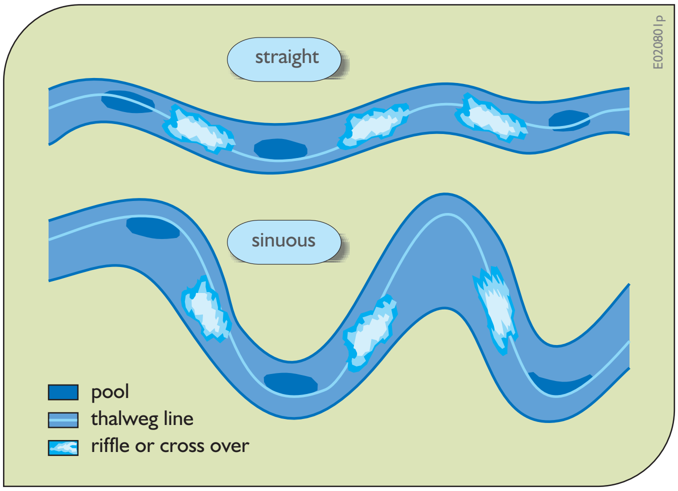



**Figure A.8.** Sequence of pools and riffles in straight and sinuous streams.

### 2.2.1. Sinuosity

Sinuosity (Figure A.8) is a term indicating the amount of curvature in the channel. The *sinuosity* of a reach is its channel centreline length divided by the length of the valley centreline. If the channel-length/valley-length ratio is more than about 1.4, the stream can be considered meandering in form. Sinuosity is generally related to streamflow power (slope times flow). Low to moderate levels of sinuosity are typically found in zones 1 and 2 of the longitudinal profile (Figure A.6). Sinuous streams often occur in the broad, flat valleys of zone 3.

### 2.2.2. Pools and Riffles

Most streams share a similar attribute of alternating, regularly spaced, deep and shallow areas called *pools* and *riffles* (Figure A.8). Pools and riffles are associated with the deepest path along the channel (thalweg). This deepest path meanders within the channel. Pools typically form in the thalweg near the outside bank of bends. Riffle areas usually form between two bends at the point where the thalweg crosses over from one side of the channel to the other. The pool-to-pool or riffle-to-riffle spacing, where they exist, is normally about five to seven times the channel width at bank-full discharge (Leopold et al., 1964).

## 2.3. Vegetation in the Stream and River Corridors

Vegetation typically varies along and across stream and river corridors. In zone 1, flood-dependent or tolerant plant communities tend to be limited but provide vegetative organic matter along with the sediment to zones 2 and 3 downstream. Woody debris from headwaters forests can be among the important features supporting food chains and instream ecological habitats in rivers (Maser and Sedell, 1994).

Zone 2 is typically a wider and more complex floodplain and larger channel than zone 1. The lower gradient, larger stream size and less steep terrain in zone 2 often allow more agricultural or residential development. This development may restrict the diversity of the natural plant communities in the middle and lower reaches, especially when land uses involve clearing and narrowing the corridor. Such actions alter stream processes involving flooding, erosion/deposition, and import or export of organic matter and sediment. This affects stream corridor geomorphology, habitat diversity, and water quantity and quality regimes.

The lower gradient, increased sediment deposition, broader floodplains and greater water volume in zone 3 typically lead to different plant communities than those found in the upstream zones. Large floodplain wetlands can develop on the flatter terrain.

The changing sequence of plant communities along streams from source to mouth is an important source of ecosystem resiliency. A continuous corridor of native plant communities is desirable. Restoring vegetative connectivity in even a portion of a stream will usually improve conditions and enhance the ecosystem's beneficial functions.

## 2.4. The River Continuum Concept

The river continuum concept identifies biological connections between the watershed, floodplain and perennial stream systems, and offers an explanation of how biological communities develop and change along perennial stream or river corridors. The concept is only a hypothesis, yet it has served as a useful conceptual model for describing some of the important features of perennial streams and rivers.

The river continuum concept assumes that, because of forest shading in many first to third-order headwater streams, the growth of algae, periphyton and other aquatic plants is limited. Since energy cannot be created through photosynthesis (autotrophic production), aquatic organisms in these lower order streams depend on materials coming from outside the channel such as leaves and twigs. Consequently, these headwater streams are considered *heterotrophic* (that is, dependent on the energy produced in the surrounding watershed). The relatively constant temperature regimes of these streams tend to limit biological species diversity.

Proceeding downstream to fourth, fifth and sixth-order streams, the channel widens, which increases the amount of incident sunlight and average temperatures. Primary production increases as a response. This shifts many stream organisms to internal autotrophic production and a dependence on materials coming from inside the channel (Minshall et al., 1985). Species richness of the invertebrate community increases due to the increase in the variety of habitats and food resources. Invertebrate functional groups, such as the grazers and collectors, increase as they adapt to both out-of-channel and in-channel sources of food.

Mid-ordered streams also experience increasing temperature fluctuations. This tends to further increase biotic diversity. Larger streams and rivers of seventh to twelfth order tend to increase in physical stability, but undergo significant changes in structure and biological function. Larger streams develop increased reliance on primary productivity by phytoplankton, but continue to receive heavy inputs of dissolved and fine organic particles from upstream.

Large streams frequently carry increased loads of clays and fine silts. These materials increase turbidity, decrease light penetration, and thus increase the significance of heterotrophic processes. The frequency and magnitude of temperature changes decrease as streamflows increase, and this in turn increases the overall physical stability of the stream as well as species competition and predation.

## 2.5. Ecological Impacts of Flow

Streamflow regimes have a major influence on the physical and biological processes that determine the structure and dynamics of stream ecosystems (Covich, 1993). High flows are important in terms of sediment transport. They also serve to reconnect floodplain wetlands to the channel. Floodplain wetlands provide habitat for aquatic plants as well as fish and waterfowl. Low flows promote fauna dispersment, thus spreading populations of species to a variety of locations. The life cycles of many riverine species require an array of different habitat types, whose temporal availability is determined by the variable flow regime. Adaptation to this environment allows riverine species to persist during periods of droughts and floods (Poff et al., 1997).

## 2.6. Geomorphology

The major large-scale physical characteristics of most streams and rivers have resulted from hydro-geological processes that have occurred over periods ranging from several decades to hundreds or even thousands of years. At smaller spatial scales, interactions between the hydrological cycle and the land that can alter the geometry of these water bodies take place in much shorter times, possibly ranging from a few minutes, hours or days to several years. Land use changes also influence the shape and flow directions of streams, rivers, estuaries and coastlines. Understanding how this happens requires a knowledge of how land cover and land-cover changes influence the partitioning of precipitation into runoff, infiltration, soil water storage, groundwater flow and storage, and discharge, and what impact all this has on the erosion, transport and deposition of sediment.

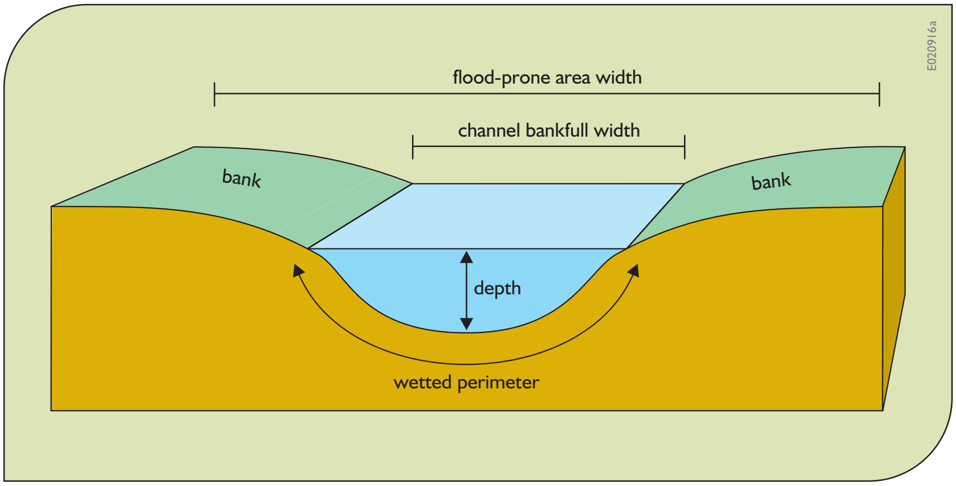



**Figure A.9. Channel cross-section schematic.**

## 2.6.1. Channel Classification

Streams and rivers are classified on the basis of their dimensions, configurations and shapes. Classifying channel morphology helps in the development of reproducible descriptions and assessments of channel flows, movement and sediment transport potential. Rosgen (1996) classified stream or river morphology based on bank-full width, channel sinuosity, slope and channel material size. While classifying a stream or river channel is helpful in understanding the behaviour of the channel and how the behaviour might change due to land use changes or modifications to the channel, it is only an aid in the management of a stream or river or in the development of an engineering design for its channel. Classification itself does not directly provide the detailed information needed for an engineering design solution.

The ratio of the flood-prone area width to the bank-full width is called the *entrenchment ratio*. The *width-to-depth ratio* is the ratio of the bank-full width to the mean bank-full depth. The *mean hydraulic depth* is the cross-sectional area (darker blue area marked as 'depth' in Figure A.9) divided by the wetted perimeter.

## 2.6.2. Channel Sediment Transport and Deposition

The energy contained in flowing water is typically expended by eroding and transporting sediment. The source of sediment comes in the runoff from the drainage area or from the channel bed and banks (Trimble and Crosson, 2000). The sediment transported in water can be dissolved, suspended and pushed along the bed by

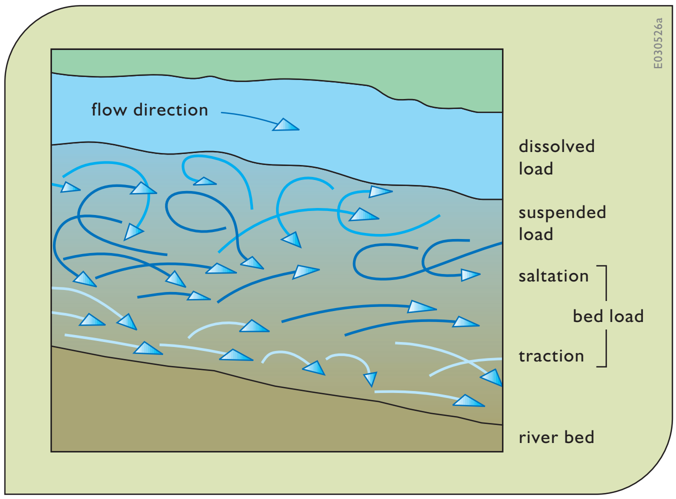


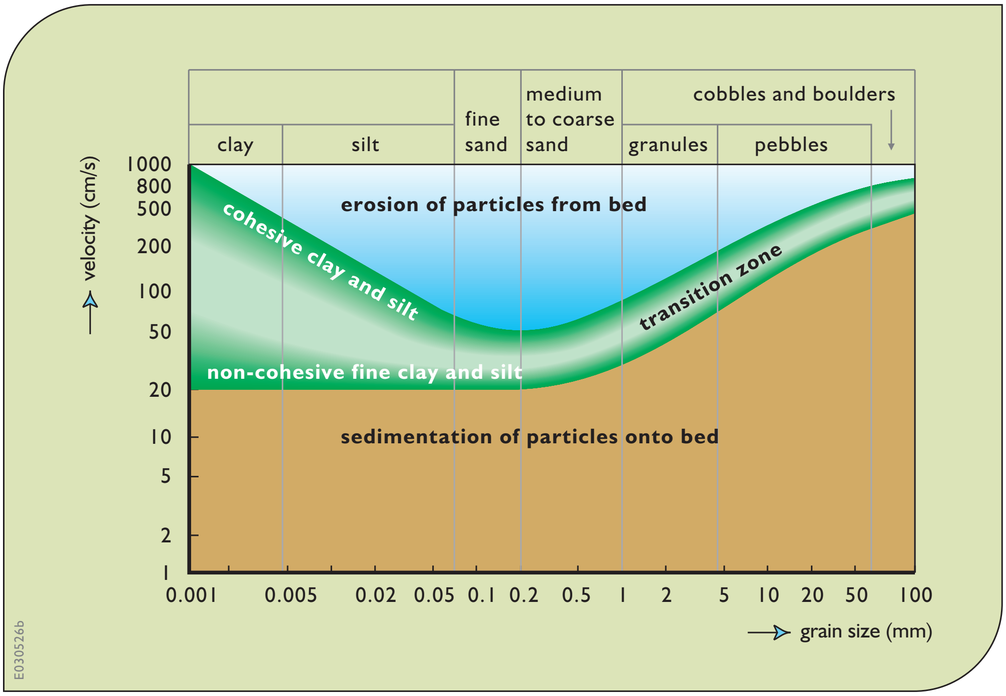

**Figure A.10. Types of sediment loads in a river channel.**

saltation and traction. The latter two processes form the bedload. Figure A.10 illustrates these types of sediment loads in a river channel.

Sediments range from clays to gravel and, in extreme events, even boulders. They include clay (<0.004 mm), silt (0.004–0.062 mm), sand (0.062–1.000 mm), gravel (granules and pebbles, 1–64 mm), cobbles (64–250 mm), and boulders (>250 mm). Each of these size classes can be subdivided from fine to coarse. In addition, fine sediments can be cohesive (tending to attach to other sediments) or non-cohesive.

Energy is required to erode and transport sediment. The heavier the sediment particles, the more energy required to erode and transport them. The energy available to do this is a function of the rate of flow. Figure A.11

Figure A.11. Relation between flow rate and sediment particle size erosion and transport.




shows the approximate flow rates, in centimetres per second, needed to transport sediment particles of various sizes. Note that the ability to erode and transport cohesive fine sediments is much less than non-cohesive sediments.

If the energy of flowing water exceeds that needed to carry the existing sediment load, additional erosion can occur to add to the sediment load. Flowing water will seek to achieve its equilibrium sediment load. Conversely if the energy is insufficient to carry the existing load in the water, some of it will be deposited.

Considerable efforts have been made over the past decades to estimate, prevent and control soil erosion and sediment runoff from watersheds. To the extent that these prevention and control efforts have worked, increased stream channel and bank erosion have occurred. Controlling watershed erosion does not reduce the sediment carrying capacity of flows. If sediment is available in or on the banks of the stream or river channel, it will replace what otherwise might have come from watershed runoff. Especially in urbanizing watersheds, where watershed contributions to sediment loads have been substantially reduced, most of the total sediment yield can be due to channel and bank erosion.

Sediment is transported as bedload and/or as suspended load. The latter consists of fine materials such as clay, silt and fine sand that usually make the flow look muddy. Flows from snow or ice melt tend to carry fine particles of rock that can make the flow look emerald green. Lake Louise in Alberta, Canada is a classic example of this. If fine materials are available, the suspended load can be as high as 95% of the total sediment load carried by the stream or river flow. This fine material settles in areas of reduced velocity. Sediment deposition processes can eventually change the channel dimensions and even its location.

Bedload transport of coarser sediments involves a combination of sliding, rolling and saltation (bouncing). These transport processes begin when the flow velocity reaches a critical value for the particular size of the bed material. This critical velocity corresponds to a critical shear stress. The shear stress associated with a flow is an important parameter for sediment transport.

Suspended sediment particle settling (or fall) velocities influence the rate of deposition. These velocities are affected by particle shapes, the density or specific weight of the sediment particles relative to that of the water, and the chemical attraction (cohesion) of the particles, especially in clays. Equations have been proposed for estimating the terminal fall velocity of a single particle in quiescent, distilled water. While equations and graphs exist in handbooks for these and other individual particle size fall velocities under quiescent distilled conditions, measurements of fall velocities under natural conditions are much more reliable.

Erosion depends on *tractive stress* or *shear stress* that creates lift and drag forces at the soil surface boundaries in fields and along the bed and banks. Shear stress varies as a function of flow depth and slope. The larger the soil particle, the greater the amount of shear stress needed to dislodge and transport it downstream. The energy differential that sets sediment particles into motion is created by faster water flowing in the main body of the channel next to slower water flowing at the boundaries. The momentum of the faster water is transmitted to the slower boundary water. The resulting shear stress moves bed particles in a rolling motion in the direction of the current.

Sediment transport rates can be computed using various models, but the uncertainty associated with any sediment load prediction can be considerable. Predicting sediment transport is one of the most enduring challenges facing hydrologists and geomorphologists.

Stream channels and their floodplains are constantly adjusting to the water and sediment in them. Channel response to changes in water and sediment yield may occur at differing times and locations, requiring differing levels of energy. Daily changes in streamflow and sediment load result in frequent adjustment of bedforms and roughness in many streams with movable beds. Streams also adjust periodically to extreme high and low-flow events. Both flood and drought flows often remove vegetation as well as creating and increasing vegetative potential along stream and river corridors. Long-term adjustments in channel structure and vegetation may come from changes in runoff or sediment yield from natural causes, such as climate change or wildfire, or from human activities such as cultivation, overgrazing or rural-to-urban conversions. Changes in vegetation can also affect streamflow and sediment deposition. Bio-geomorphology is the term used for the study of these plant, soil and water interactions.

### 2.6.3. Channel Geometry

Like all physical systems, stream and river flows and their sediment loads will attempt to reach an equilibrium. The equilibrium between flows and sediment erosion, uptake and deposition, is sometimes referred to as *regime theory*. Leopold and Maddock (1953) derived regime relations in stream and river channels in alluvial (sedimentary) basins. These relations predicted the stream width, $w$; mean hydraulic depth, $h$; velocity, $U$; sediment concentration, $Q_s$; slope, $S$; and Manning's $n$ friction coefficient as a power function of the discharge, $Q$, in the stream or river. Each of these relations $i$ is of the form $a_iQ^{bi}$. Leopold and Maddock found the values of these different coefficients $a_i$ and $b_i$ for a variety of rivers in the United States (also in Richardson et al., 1990).

The values of the coefficients $a_i$ and $b_i$ may differ along the length of the stream or river (Leopold, 1994). In the downstream direction the relative change in the width will increase more rapidly than the relative change in velocity. Hence, the coefficients $a_i$ and $b_i$ associated with width, depth and velocity will change. However, since the product of width, depth and velocity equals the flow $Q$, and each equation $a_iQ^{bi}$ for width, depth and velocity are functions of the flow $Q$, both the product of the three coefficients $a_i$ and the sum of the three exponent coefficients $b_i$ must equal 1.

Other studies have found approximate relationships between bank-full channel dimensions of alluvial streams and rivers and their bank-full discharge. The channel dimensions, pattern and profile are primarily related to the effective or bank-full discharge. The magnitude of the bank-full discharge in the main channel typically corresponds to the 1.5- to 2-year expected return period flow event based on annual peak flows. The bank-full stage is lower than the top of the bank. It is commonly identified as a bench, a change in bank material and vegetation, or the top of a point bar (Rosgen, 1996; Ward and Elliot, 1995).

The log–log plot of Figure A.12 shows a relationship between flow rates and depths in relation to bank-full flow rates and depths for thirteen rivers in the eastern United States (Leopold et al., 1992).

The dimensionless rating curve presented in Figure A.12 shows that the annual discharge is less than the bank-full

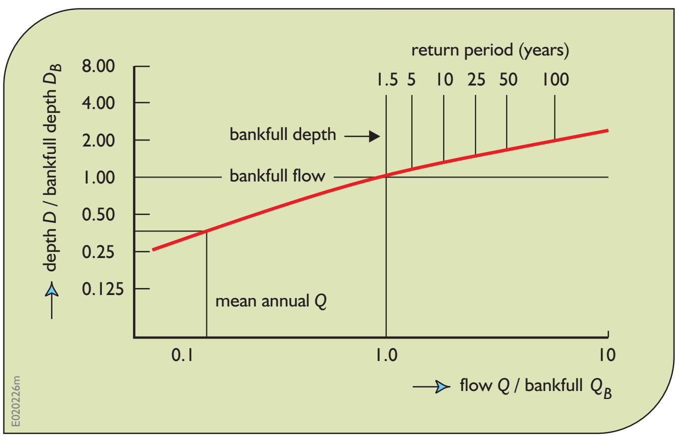



**Figure A.12.** Dimensionless regression curve based on thirteen gauging stations in the eastern United States (Leopold, Wolman and Miller, 1992).

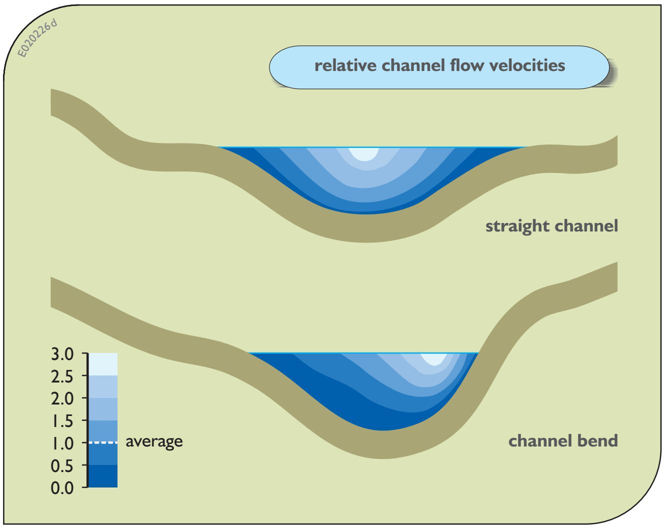



**Figure A.13. Typical channel relative velocities at bank-full flows. The velocities at the outside of channel bends (the right-hand side of the lower channel) are greater than on the inside of the bends.**

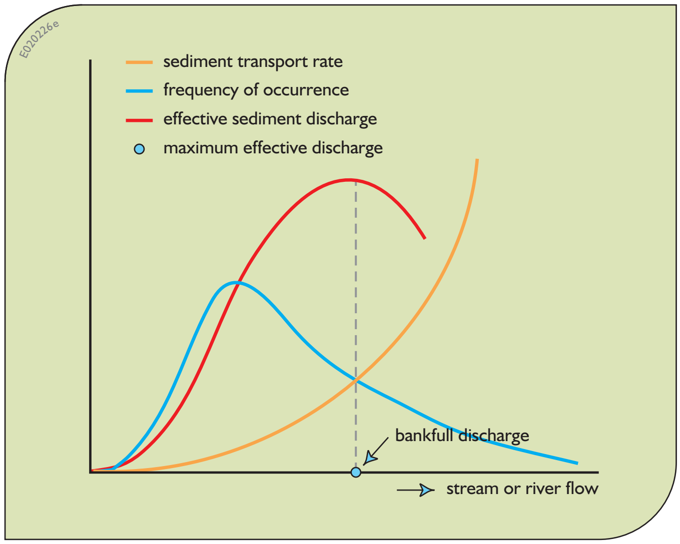



**Figure A.14. The bank-full discharge typically has the largest total sediment load.**

discharge and occurs at a stage of about a third of the bank-full depth. A 50- to 100-year return period discharge occurs at a stage that is about double the maximum bank-full depth. In developing a stream classification system, Rosgen (1996) defined a flood prone width as one that occurs at twice the maximum bank-full depth. Therefore, it might be expected that this depth corresponds to about the 100-year return period discharge.

### 2.6.4. Channel Cross Sections and Flow Velocities

Typically, flow velocity across a channel will vary from one bank to another. This lateral variation will change along a stream or river channel. For example, the velocity near the outer bank of a meander will be higher than near the inner bank because water near the outer bank has to travel further. Also, the velocity decreases towards the bottom or sides of the channel due to the surface roughness of the channel. Water in contact with the bottom and sides will be stationary. These relationships are shown in Figure A.13.

The velocity distribution in a channel can change from one cross section to the next. Typically, at any cross section in a channel, there will be a portion of the flow above the deepest point (the thalweg) that is moving the fastest. Along the thalweg the velocity increases in riffles, decreases in pools, moves from side to side, up and down with depth, and rotates as it moves around bends. The rotation of the flow on the outer bend might also create undercut concave banks.

The velocity decreases near the bottom or sides of the channel due to the roughness of its surface. Water in contact with the bottom and sides will be stationary. The changing velocities from riffles to pools will cause turbulence. Also, as water moves through pools, it will push towards the banks and away from its direction of flow. As much as a third of the flow might circulate back upstream.

The bank-full discharge is most effective in forming a channel, benches (active floodplains), banks and bars. It is this discharge rate that, as shown in Figure A.14, transports the largest total sediment load over time. The sediment concentration in the water is a function of the flow discharge rate and of course the available sediment. Low discharge rates are ineffective in transporting sediment, while high ones have very high sediment transport rates. However, extremely high discharge events occur less frequently, so the total sediment load they carry over a period of many years is not the largest.

A measure of the total sediment load carried by a particular discharge equals the event frequency multiplied by the transport rate for that discharge. The maximum value of that product for all discharges is called the *effective discharge*. This effective discharge is the bank-full discharge. It has the longest-term impact on an alluvial stream's or river's equilibrium morphology.

Discharges larger than the mean annual discharge typically transport more than 95% of the sediment in a river system.

### 2.6.5. Channel Bed Forms

Bed forms result from the interaction of the fluid forces and sediment particles. Bed forms in alluvial streams and rivers also depend on channel shape, size of bed material, bed vegetation and the viscosity of the fluid. In flumes having a limited range in depth and discharge, changing the slope is the principal means of changing the bed forms. In streams and rivers, however, where the slope is relatively constant, a change in bed form can occur with a change in discharge and/or a change in viscosity. Viscosity is affected by a change in sediment size distribution and/or temperature. An increase in viscosity resulting from a decrease in temperature or an increase in fine sediments such as clays can change a dune bed to washed-out dunes, plane beds or antidunes. For example, the Missouri River in the United States along the border between Iowa and Nebraska has temperatures between 21 and 27 °C in the summer and between 15 and 17 °C in the autumn. In summer, its bed form is characterized by dunes. In autumn, its bed form becomes washed-out dunes (USACE, 1969).

### 2.6.6. Channel Planforms

Channel planforms are what one sees when looking at them from above. Their geographic shapes can be broadly grouped into straight, meandering and braided planforms (Lagasse et al., 1991; Leopold and Wolman, 1957; Schumm, 1972, 1977; Richardson et al., 1990). These three types of channels are shown in Figure A.15. More detailed classifications have been used (Brice and Blodgett, 1978; Culbertson et al., 1967), but these three basic types are the ones most commonly considered.

A meandering stream is characterized by sinuous S-shaped flow patterns. The sinuosity of a channel of fixed

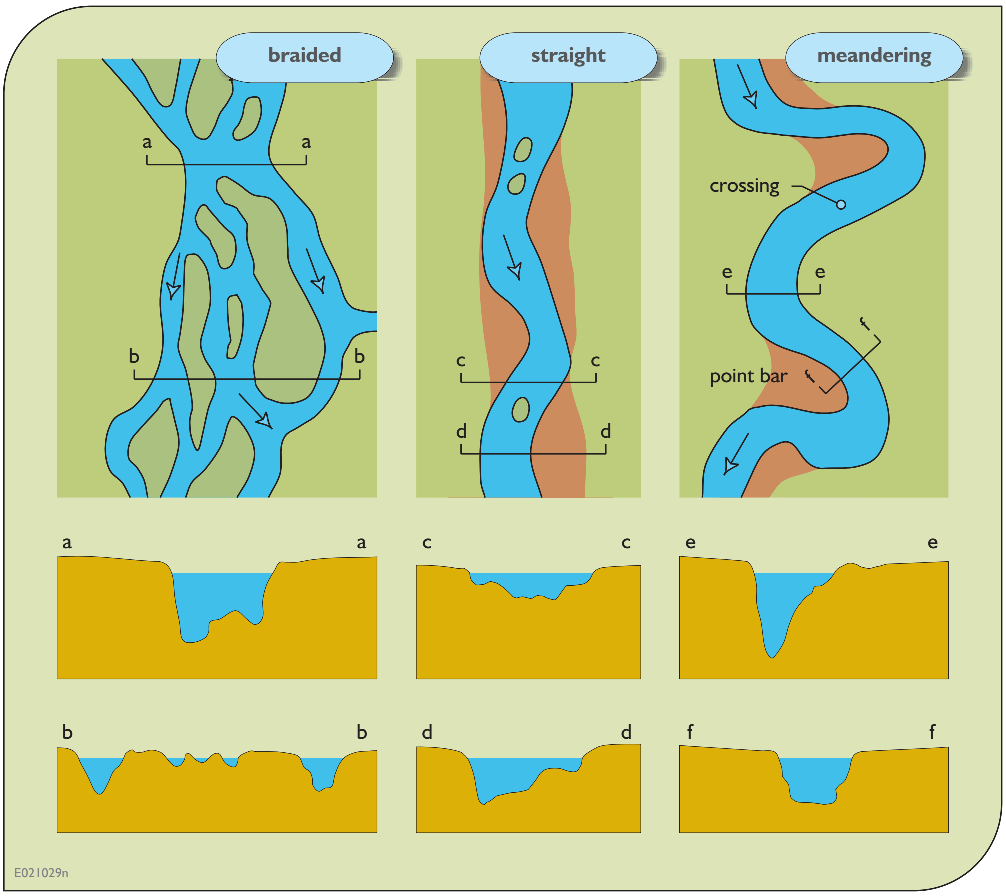



**Figure A.15.** Types of channel planforms (after Richardson et al., 1990).

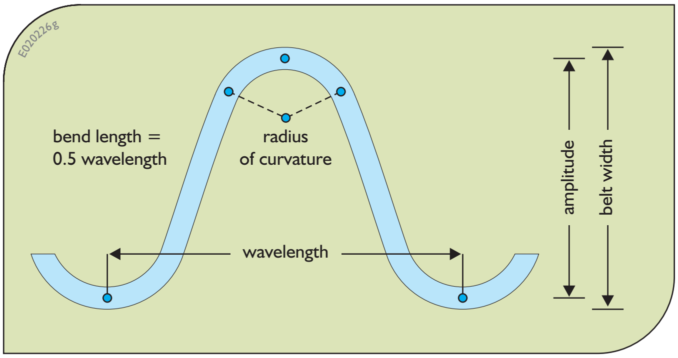



**Figure A.16. Geometry of a meandering stream (Rosgen, 1996).**

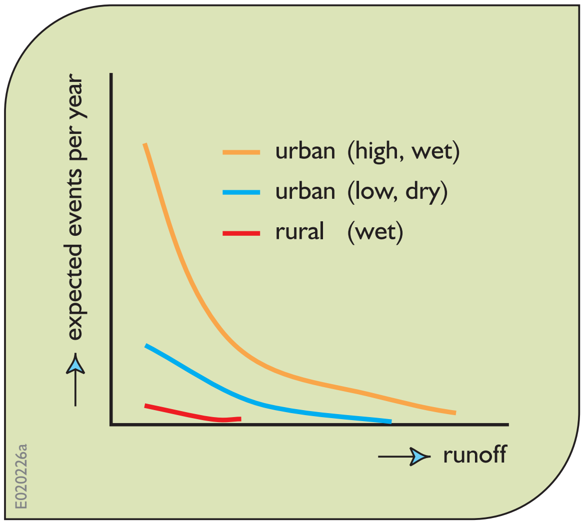



**Figure A.17. Expected annual occurrences of runoff events showing the effect of high and low urban and rural development densities under wet and dry soil moisture conditions at the beginning of a storm (Ward, personal communication).**

length is the ratio between its length and the straight-line distance between channel end points, or valley length. Channels can be classified according to their sinuosity. If the sinuosity is 1.0 the channel is called straight, a condition that rarely occurs in natural alluvial channels except over short distances. If the sinuosity is greater than 1.0 the channel is called sinuous. If the sinuosity is greater than about 1.4 the river is said to meander, and if the sinuosity exceeds 2.1 the degree of meandering is tortuous.

Figure A.16 defines some geometric parameters of meandering channels.

A braided stream or channel consists of a number of subordinate channels. At normal and low flow rates, these subordinate channels are separated by bars, sandbars or islands. The shape and location of the bars and islands can change with time, sometimes with each runoff event. In an anabraided stream or river, the subordinate channels are more permanent and more widely and distinctly separated than those of a braided stream.

A stream or river channel can have different planforms at different locations along its length. Channel planforms can affect channel dimensions, flows, bed material, floodplain size and plant cover. Channel planforms in turn are affected by geology, topography, the drainage area of the contributing watershed, flow velocity, discharge, sediment transport, sediment particle distribution, channel geometry, vegetation cover and any geomorphologic controls on the system.

### 2.6.7. Anthropogenic Factors

Stream and river morphology is affected by human activities such as changing land use, bank protection, navigation, and construction and operation of hydraulic infrastructure. Deforestation and cultivation of watershed areas, water management initiatives and floodplain development affect the runoff and the sediment yield to the river; they also influence the hydraulic roughness and the trapping of sediment during floods.

Urbanization increases the fraction of impervious land in a drainage area. This in turn reduces infiltration and causes more runoff and higher peak discharges. Sediment loads typically increase during construction and decrease following construction. All of these factors affect the equilibrium in the river and can cause it to widen and/or deepen. Channel modifications such as straightening a reach of the channel will also increase the sediment-carrying capacity of the discharge and might cause bank and bed scour.

Urbanization increases the frequency as well as the amounts of water associated with different runoff events (Figure A.17). Increasing imperviousness increases the number of expected runoff events each year. The soil moisture that exists at the beginning of a storm also influences the frequency of runoff events. In areas of Ohio in the United States, Ward (personal communication) found at the start of a storm event in a rural area with wet soil conditions there might be only one 1-cm runoff event annually and a 2-cm runoff event every few years. However, a low-density urban area or an urban area with dry soil conditions at the start of a storm event might on average experience six or seven 1-cm runoff events, two or three 2-cm events, one 2.5-cm event per year, and even an event with 3.5-cm every few years. High-density urban areas, or urban areas with wet soil conditions at the start of a storm event will on average experience more than twenty 1-cm runoff events, six 2-cm events, two 2.5-cm events, and one 3.5-cm event per year. These were for this study area in Ohio; for other areas elsewhere the numbers will differ, but the general relationships shown in Figure A.17 will apply.

The effect of urbanization on discharges is also illustrated in Figures A.18 and A.19. Rural areas will have little if any impervious cover, while urban areas will have considerably more. The plot shows the relative discharge (percentage of the rural discharge) increase that occurs as the percentage of impervious cover increases. Plots are presented for flow return periods of 2, 10 and 100 years. Note that the biggest impact is on the smaller, more frequent storm events. The two-year return period discharges, $Q_2$, in highly urbanized areas can be over two-and-a-half times larger than those on rural areas. In addition, as imperviousness increases, so does channel width.

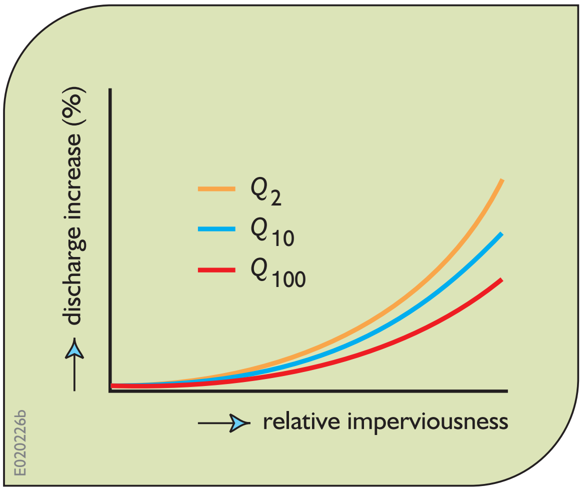



**Figure A.18.** Relative discharge increases, for flows having 2- 10- and 100- year return periods, as a function of degree of impervious surface area.

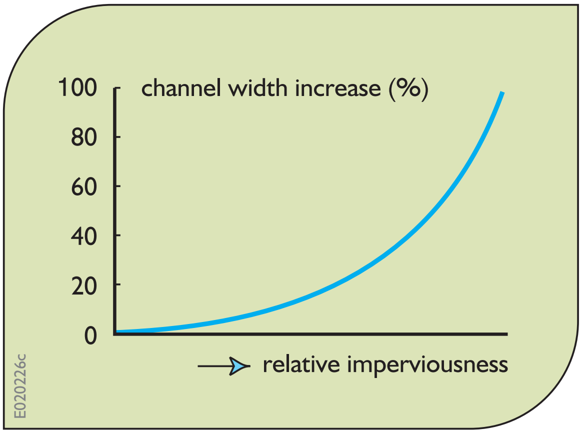



**Figure A.19.** Relative increase in channel width as a function of impervious surface area.

## 2.7. Water Quality

Establishing an appropriate flow regime in a stream or river corridor may do little to ensure a healthy ecosystem if the physical and chemical characteristics of the water are damaging to that ecosystem. For example, streams or rivers with high concentrations of toxic materials, high temperatures, low dissolved oxygen concentrations or other harmful physical/chemical characteristics cannot support healthy stream corridor ecosystems.

Figure A.20 illustrates some of the key water quality processes affecting the oxygen content and hence the biology of surface waters. (The modelling of these and other chemical and biological processes is discussed in Chapter 12.)

Dissolved oxygen (DO) is a basic requirement for a healthy aquatic ecosystem. Most fish and aquatic insects 'breathe' the oxygen dissolved in the water body. Some fish and aquatic organisms, such as carp and sludge worms, are adapted to low oxygen conditions, but most sport-fish species, such as trout and salmon, suffer when DO concentrations fall below a concentration of 3–4 mg/l. Larvae and juvenile fish are more sensitive and require even higher concentrations of DO.

Many fish and other aquatic organisms can recover from short periods of low oxygen concentrations in the water. However, prolonged episodes of depressed dissolved oxygen concentrations of 2 mg/l or less can result in 'dead' (anaerobic) water bodies. Prolonged exposure to low DO conditions can suffocate adult fish or reduce their reproductive survival rate by suffocating sensitive eggs and larvae, or starve fish by killing aquatic insect larvae and other sources of food. Low DO concentrations also favour anaerobic bacteria that produce the noxious gases often associated with polluted water bodies.

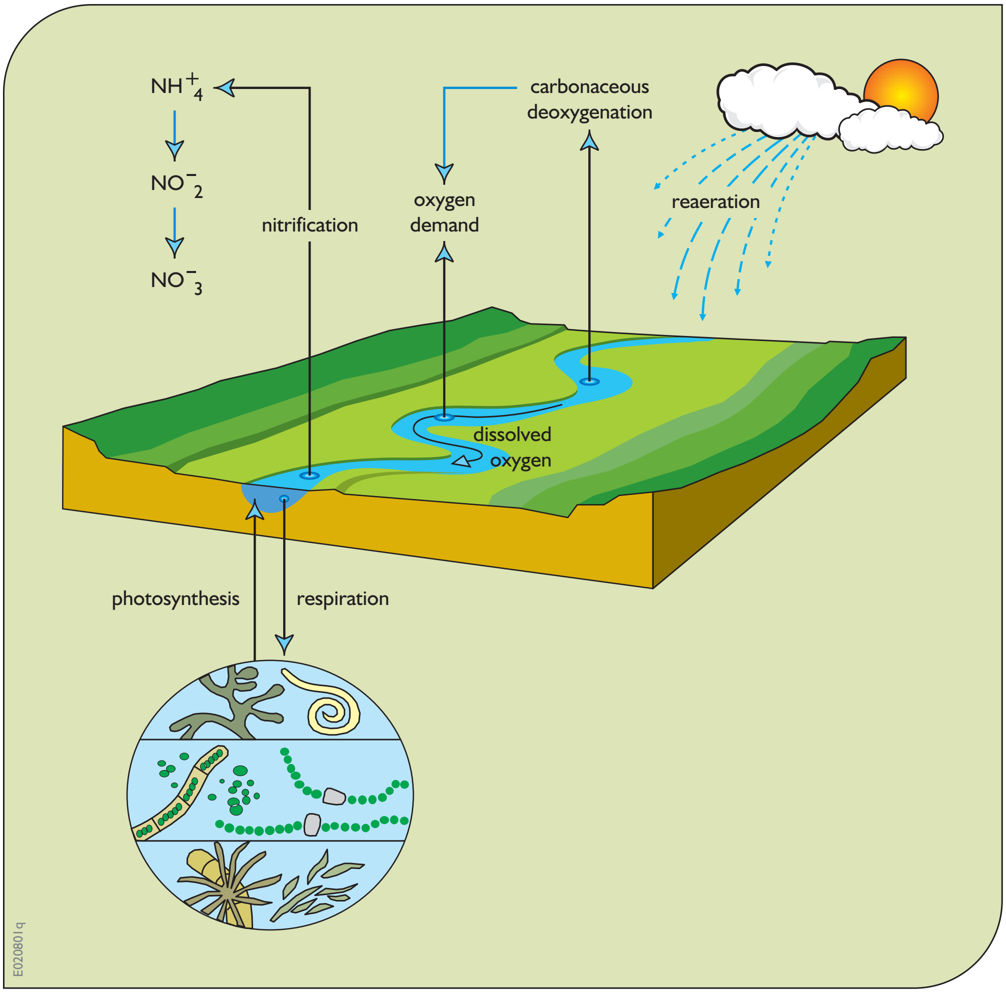



**Figure A.20.** Some of the basic chemical and biological processes affecting dissolved oxygen in waters.

Water absorbs oxygen directly from the atmosphere, and from plants as a result of photosynthesis. The amount of oxygen that can be dissolved in water is influenced by its temperature and salinity. Water loses oxygen primarily by respiration of aquatic plants and animals, and by the mineralization of organic matter by microorganisms. Discharges of oxygen-demanding wastes or excessive plant growth (eutrophication) induced by nutrient loading followed by death and decomposition of vegetative material can also deplete oxygen.

In addition to oxygen and water, aquatic plants require a variety of other elements to support their bodily structures and metabolism. Just as with terrestrial plants, nitrogen and phosphorus are important among these elements. Additional nutrients, such as potassium, iron, selenium and silica, are also needed by many species but are generally not limiting factors to plant growth. When any of these elements are limited, plant growth may be limited. This is an important consideration in ecosystem management.

Nutrients cycle from one form to another depending on nutrient inputs, as well as temperature and available oxygen. The nitrogen cycle is illustrated in Figure A.21 as an example. Table A.1 lists some common sources of nitrogen and phosphorus nutrient inputs and their typical concentration ranges.

Management activities can interact in a variety of complex ways with water quality. This in turn can affect ecosystem species, as shown in Figure A.22.

## 2.8. Aquatic Vegetation and Fauna

Stream biota are often classified in seven groups: bacteria, algae, macrophytes (higher plants), protists (amoebas, flagellates, ciliates), microinvertebrates (such as rotifers,

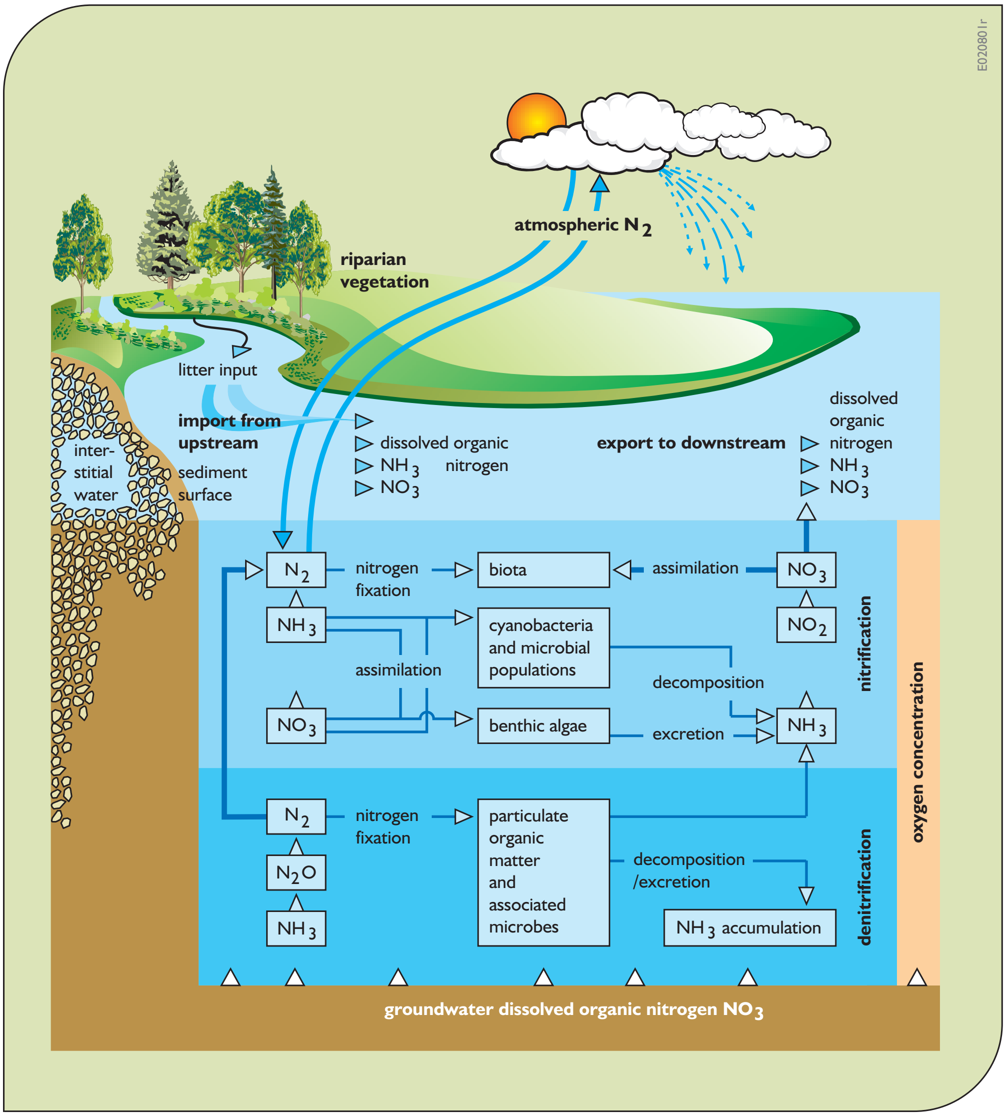



**Figure A.21.** Dynamics and transformations of nitrogen in a stream ecosystem.

E020903e

| source | total nitrogen (mg/l) | total phosphorus (mg/l) |
|---|---|---|
| urban runoff | 3-10 | 0.2-1.7 |
| livestock operations | 6-800 | 4-5 |
| atmosphere (wet deposition) | 0.9 | 0.015 |
| 90 % forest | 0.06-0.19 | 0.006-0.012 |
| 50 % forest | 0.18-0.34 | 0.013-0.015 |
| 90 % agriculture | 0.77-5.04 | 0.085-0.104 |
| untreated waste water | 35 | 10 |
| treated waste water | 30 | 10 |

**Table A.1.** Sources and concentrations of pollutants from common point and non-point sources (USDA, 1998). These data show little or no impact on nutrient removals from basic wastewater treatment facilities.

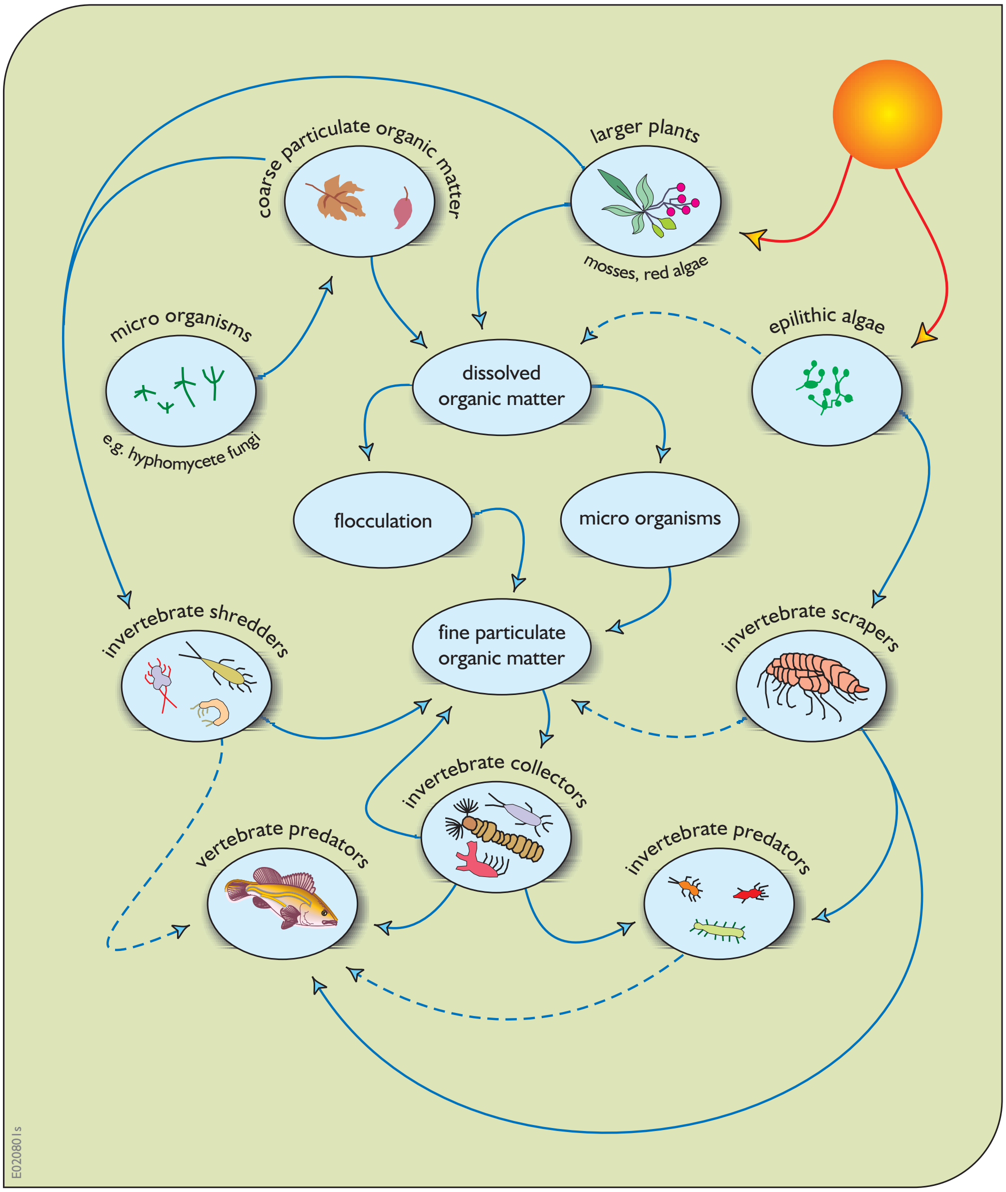



**Figure A.22.** Interactions among aquatic organisms and their sources of energy. (Dashed lines reflect weaker interactions.)

copepods, ostracods, and nematodes), macroinvertebrates (such as mayflies, stoneflies, caddisflies, crayfish, worms, clams, and snails) and vertebrates (fish, amphibians, reptiles, and mammals). The river continuum concept provides a framework for describing how these organisms change from lower to higher-order streams.

Much of the spatial and temporal variability of type, growth, survival and reproduction of aquatic organisms reflects variations in water quality, temperature, streamflow and flow velocity, substrate content, the availability of food and nutrients, and predator–prey relationships. These factors are often interdependent.

## 2.9. Ecological Connectivity and Width

Healthy ecosystems also depend on conductivity and width. Connectivity is a measure of how spatially continuous a corridor or a matrix is (Forman and Godron, 1986). A stream corridor with connections among its natural communities promotes transport of materials and energy and movement of flora and fauna. Connectivity is illustrated in Figure A.23.

Width is the distance across the stream and its zone of adjacent vegetation cover. Factors affecting width are edges, community composition, environmental gradients

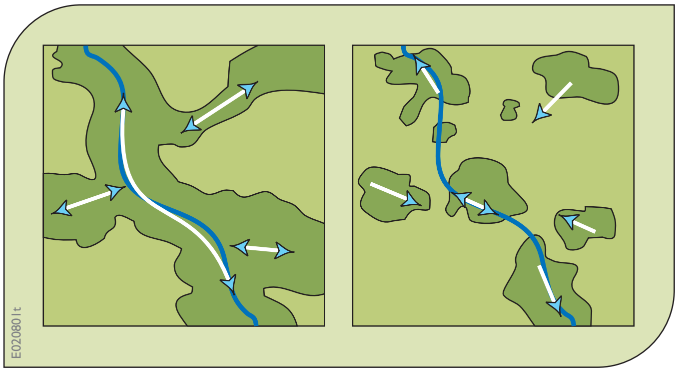



**Figure A.23.** Landscapes with high (left picture) and low (right picture) degrees of connectivity.

and disturbance effects of adjacent ecosystems, including those due to human activity. Width and connectivity interact throughout the length of a stream corridor. Corridor width varies along a stream and may have gaps. Gaps across the corridor can interrupt and reduce connectivity. Ensuring connectivity and adequate width can provide some of the most useful ways to mitigate disturbances.

## 2.10. Dynamic Equilibrium

In constantly changing ecosystems like stream and river corridors, the stability of a system is its ability to persist within a range of conditions. Within this range the system is resilient. This phenomenon is referred to as *dynamic equilibrium*.

The maintenance of dynamic equilibrium requires a series of self-correcting mechanisms in the stream corridor ecosystem. These mechanisms control the responses to external stresses or disturbances within certain ranges. The threshold levels associated with these ranges are often difficult to identify and quantify. If they are exceeded, the system can become unstable. Corridors may then undergo a change toward a new steady-state condition, usually after some time for readjustment has occurred.

Many stream systems can accommodate some disturbances and still return to functional conditions once the sources of the disturbances are removed. Ecosystems tend to heal themselves when external stresses are removed. The time it takes to do this, of course, depends on the level of stress.

## 2.11. Restoring Degraded Aquatic Systems

Some principles found useful in the management of restoration activities in degraded stream or river systems are listed below. These principles apply from early planning to post-implementation monitoring. They focus on scientific and technical measures and their likely impacts. Restoration activities can include upland best-management practices for agriculture, stream channel restoration, removal of exotic species and increasing native plant cover, establishing windbreaks and shelterbelts, improving upland corridors and riparian habitat, and wetland enhancement for water quality improvements. As in all management activities, the presence or absence of public support for a restoration project can make the difference between success and failure.

### Preserve and Protect Aquatic Resources

Existing ecosystems that are relatively intact provide the natural materials needed for the recovery of impaired systems. It is not usually necessary to import species.

### Restore Ecological Integrity

Restoration should re-establish the ecological integrity of degraded aquatic ecosystems. Ecological integrity refers to the natural structure, composition and processes of biotic communities and the physical environment. Key processes, such as nutrient cycles, succession, water levels and flow patterns, and the dynamics of sediment erosion and deposition, function within the natural range of variability. Restoration strives toward ecological integrity by taking actions that favour the desired natural processes and communities that can be sustained through time.

### Restore Natural Structure and Function

Many aquatic resources in need of restoration have problems caused by past changes in channel form or other physical characteristics that led to problems such as habitat degradation, changes in flow regimes and siltation. Stream channelization, ditching in wetlands, disconnection from adjacent ecosystems and shoreline modifications are examples of such adverse changes.

Structure and function are closely linked in river corridors, lakes, wetlands, estuaries and other aquatic resources. Re-establishing the appropriate natural structure can bring back beneficial functions. For example, restoring the bottom elevation in a wetland can help re-establish the hydrological regime, natural disturbance cycles and nutrient fluxes. Monitoring the extent to which desired functions have been re-established can be a good way to determine the effectiveness of a restoration project.

### Work Within the Watershed and Broader Landscape Context

Restoration requires a design based on the entire watershed or river basin, not just the part of it that is degraded. A localized restoration project may not be able to change what goes on in the whole watershed, but it can be designed to accommodate watershed effects better. New and future urban development may, for example, increase runoff volumes, streambed down-cutting, bank erosion and pollutant loading. Restoration may help mitigate these adverse effects. For example, in choosing a site for a wetland, stream or river restoration project, planners should consider how the proposed project may be used to further related efforts in the watershed, such as increasing riparian habitat continuity, reducing flooding and/or enhancing downstream water quality. Beyond the watershed, the broader landscape context also influences restoration through factors such as interactions with terrestrial habitats in adjacent watersheds, or the deposition of airborne pollutants from other regions.

### Understand the Natural Potential of the Watershed

A watershed has the capacity to become only what its physical and biological setting – its climate, geology, hydrology and biological characteristics – will support. Establishing restoration goals for a water body requires knowledge of the natural range of conditions that existed on the site prior to degradation and of what future conditions might be. This information can then be used in determining appropriate goals for the restoration project.

### Address Ongoing Causes of Degradation

Restoration efforts are likely to fail if the causes of degradation persist. Therefore, it is essential to identify the causes of degradation and eliminate or mitigate them wherever possible. Degradation can be caused by one event, such as the filling of a wetland, or it can be caused by the cumulative effect of numerous events, such as gradual increases in the amount of impervious surfaces in the watershed that alter the streamflow regime. In identifying the sources of degradation, it is important to look at upstream and upslope activities as well as at direct impacts on the immediate site where damage is evident. In some situations, it may also be necessary to consider downstream modifications such as dams and channelization.

### Develop Clear, Achievable and Measurable Goals

Goals direct implementation and provide the standards for measuring success. The chosen goals should be achievable given the natural potential of the area and the available resources and the extent of community support for restoration. Goals provide focus and increase project efficiency. They can also change (adapt) over time.

### Focus on Feasibility

Particularly in the planning stage, it is critical to focus on whether any proposed restoration activity is feasible, taking into account scientific, financial, social and other considerations. Community support for a project is needed to ensure its long-term viability. Ecological feasibility is also critical. For example, a wetland, stream or river restoration project is not likely to succeed if the hydrological regime that existed prior to degradation cannot be re-established.

### Use a Reference Site

Reference sites are areas that are comparable in structure and function to the proposed restoration site before it was degraded. They may be used to identify targets for restoration projects, and as yardsticks for measuring the progress of the project. While it is possible to use historic information on sites that have been altered or destroyed, it may be most useful to identify an existing, relatively healthy, similar site as a benchmark.

### Anticipate Future Changes

Although it is impossible to plan for the future precisely, foreseeable ecological and societal changes can and should be factored into restoration design. For example, in repairing a stream channel, it is important to take into account potential changes in runoff resulting from projected increases in upstream impervious surface area due to development. In addition to potential impacts from changes in watershed land use, natural changes such as plant community succession can also influence restoration. Long-term, post-project monitoring should take into account successional processes in a stream corridor when evaluating the outcome of the restoration project.

### Involve the Skills and Insights of a Multidisciplinary Team

Restoration can be a complex undertaking that integrates a wide range of disciplines, including ecology, aquatic biology, hydrology and hydraulics, geomorphology, engineering, planning, communications and social science. The planning and implementation of a restoration project should involve people with experience in the disciplines needed for that particular scheme. Complex restoration projects require effective leadership to bring the various disciplines, viewpoints and styles together as an effective team.

### Design for Self-Sustainability

Perhaps the best way to ensure the long-term viability of a restored area is to minimize the need for continuous operation, maintenance and repair costs, vegetation management or frequent repair of damage done by high-water events. High-maintenance approaches make long-term success dependent upon human and financial resources that may not always be available. In addition to limiting the need for maintenance, designing for self-sustainability also involves favouring ecological integrity. An ecosystem in good condition is more likely to have the ability to adapt to changes.

### Use Passive Restoration, When Appropriate

Before actively altering a restoration site, determine whether simply reducing or eliminating the sources of degradation and allowing time for recovery will be enough to allow the site to regenerate naturally. There are often reasons for restoring a water body as quickly as possible, but there are other situations when immediate results are not critical. For some rivers and streams, restoring the original hydrological regime may be enough to let time re-establish the native plant community, with its associated habitat value.

### Restore Native Species and Avoid Non-Native Species

Many invasive species out-compete native species because they are expert colonizers of disturbed areas and lack natural controls. The temporary disturbance present during restoration projects invites colonization by invasive species that, once established, can undermine restoration efforts and lead to further spread of these invasive species. Special attention should be given to avoiding the unintentional introduction of non-native species at the restoration site when the site is most vulnerable to invasion. In some cases, removal of non-native species may be among the primary goals of the restoration project.

### Use Natural Fixes and Bioengineering Techniques, Where Possible

Bioengineering is a method of construction combining live and dead plants or inorganic materials, to produce living, functioning systems to prevent erosion, control sediment and other pollutants, and provide habitat. Bioengineering techniques can often be successful for erosion control and bank stabilization, flood mitigation and even water treatment. Specific projects can range from the creation of wetland systems for the treatment of stormwater to the restoration of vegetation on riverbanks to enhance natural decontamination of runoff before it enters the river.

### Monitor and Adapt Where Changes are Necessary

Every combination of watershed characteristics, sources of stress, and restoration techniques is unique and, therefore, restoration efforts may not proceed exactly as planned. Adapting a project to at least some change or new information should be considered normal. Monitoring before and during the work is crucial for finding out whether goals are being achieved. If they are not, adjustments should be undertaken. Post-project monitoring will help determine whether additional actions or adjustments are needed, and can provide useful information for future restoration efforts. This process of monitoring and adjustment is known as adaptive implementation or adaptive management (Appendix B). Monitoring plans should be feasible in terms of costs and technology, and should provide information relevant to meeting the project goals.

# 3. Lakes and Reservoirs

Lakes and reservoirs are components of many river systems. They are typically dramatic and visually pleasing features of a watershed or basin. They range from pond-sized water bodies to lakes stretching for hundreds of kilometres. Referred to by some as 'pearls on a river', lakes and reservoirs can have a significant effect on the quantity and quality of the freshwater that eventually reaches the oceans.

Seen from the shoreline, a large natural lake looks much the same as a large artificial reservoir, and both often contain the word 'lake' in their name. Furthermore, the same principles of biology, chemistry and physics apply to both. Indeed, it may be difficult to discern any obvious differences between a lake and reservoir, but differences as well as similarities exist.

Lakes are water bodies formed by nature whereas reservoirs are artificial ones constructed by humans, either by damming a flowing river or by diverting water from a river to an artificial basin (impoundment). Some reservoirs are made by increasing the capacity of natural lakes. Many characteristics of lakes and reservoirs are a function of the way in which they were formed and how humans use their waters.

Lakes and reservoirs are important sources of freshwater for agriculture, industry and municipalities. At the same time, they provide habitats for a variety of species of plants and animals. They are the sources of fish, areas for migratory birds to feed, reproduce or rest, and places we all go for enjoyment and recreation. Considerable money as well as technical and scientific expertise is often required to keep them clean and healthy.

As human populations grow, greater demands are placed on the services lakes and reservoirs provide. The water levels of many lakes and reservoirs have become consistently lower as a result of higher consumption by upstream agriculture, households and industries. Increasing numbers of people enjoy the recreational activities these bodies can support, but this alters their shores, changes the surrounding land use and cover, and increases the amount of soil or sediments as well as nutrients reaching their waters. Finally, pollution from adjacent lands and various point sources may increase eutrophication processes and produce other non-desirable effects such as increased concentrations of toxic algae, reduced dissolved oxygen and generation of foul odours.

## 3.1. Natural Lakes

Lakes are naturally-formed, usually bowl-shaped, depressions in the land surface that have filled with water over time. These depressions were typically produced as a result of glaciers, volcanic activity or tectonic movements. The age of most permanent lakes is usually of a geological time frame. Some ancient lakes may be millions of years old.

The most significant past mechanism for the formation of lakes in temperate areas was the natural process of 'glacial scour', in which the slow movement of massive volumes of glacial ice during and after the Ice Age produced depressions in the land surface that subsequently filled with water. The North American Great Lakes (Superior, Michigan, Huron, Erie and Ontario), lakes in the Lake District of the United Kingdom, and the numerous lakes in Scandinavia and Argentina are prominent examples of this type of lake formation. Some smaller 'kettle lakes', as found on Cape Cod in Massachusetts, for example, were formed by the deposition and subsequent melting of glacial ice blocks.

Another major lake formation process was 'tectonic movement', in which slow movements of the earth's crust gradually produced depressions that were subsequently filled with water. Lake basins also formed as a result of volcanic activity, which also produced depressions in the land surface. Most of the earth's very deep lakes resulted from either volcanic or tectonic activity. Lake Baikal in Russia, the world's deepest lake, which contains approximately 20% of the world's liquid freshwater, and the African Rift Valley lakes are prominent examples of this tectonic type of lake formation.

Other natural processes that produced lake basins include seepage of water down through layers of soluble rock, erosion of the land surface by wind action, and plant growth or animal activity (such as beaver dams) that resulted in blocking the outlet channels from shallow depressions in the land surface.

There are literally millions of small lakes around the world, concentrated largely in the temperate and sub-arctic regions. These regions are also characterized by a relative abundance of freshwater. Many more intermittent lakes occur in semi-arid and arid regions.

## 3.2. Constructed Reservoirs

In contrast to natural processes of lake formation, reservoirs are water bodies that are usually formed by constructing a dam across a flowing river. A dam may sometimes also be constructed on the outlet channel of a natural lake as a means of providing better control of the lake's water level (examples include Lake Victoria in Africa, and Lake Tahoe in the United States). However, these latter water bodies typically retain their natural lake characteristics.

Reservoirs are found primarily in areas with relatively few natural lakes, or where the lakes do not satisfy human water needs. Reservoirs are much younger than lakes, with life spans expressed in terms of historical rather than geological time. Although lakes are used for many of the same purposes as reservoirs, a distinct feature of the latter is that they are usually built to address specific water needs. These include municipal and drinking water supplies, agricultural irrigation, industrial and cooling water supplies, power generation, flood control, sport or commercial fisheries, recreation, aesthetics and/or navigation. Small reservoirs are sometimes built for fire protection as well.

The reasons for constructing reservoirs are ancient in origin, and have focused on the need of humans to ensure a more reliable water supply and to protect themselves during periods of floods. Accordingly, reservoirs are usually found in areas of water scarcity, or where a controlled water facility was necessary. Small reservoirs were first constructed some 4,000 years ago in China, Egypt and Mesopotamia, primarily to supply drinking water and for irrigation purposes. Simple small dams were constructed by blocking a stream with soil and brush, in much the same manner as beavers dam a stream. Larger reservoirs were constructed by damming a natural depression, or by forming a depression along the river and digging a channel to divert water to it from the river. Early irrigation practices were linked largely to land adjacent to streams. They required the construction of larger dams that allowed humans to impound larger volumes of water. Later reservoirs were also used as sources of power, first to drive waterwheels and subsequently to produce hydroelectric power.

## 3.3. Physical Characteristics

Like lakes, reservoirs range in size from pond-like to very large water bodies. The variations in type and shape, however, are much greater than for lakes. The term 'reservoir' includes different types of constructed water bodies and/or water storage facilities. These are (1) valley reservoirs, created by constructing a barrier (dam) perpendicular to a flowing river; and (2) off-river storage reservoirs, created by constructing an enclosure parallel to a river, and subsequently supplying it with water either by gravity or by pumping from the river. The latter reservoirs are sometimes also called embankment or bounded or pumped-storage reservoirs. They have controlled inflows and outflows to and from one or more rivers. For example, much of the water in the river above Niagara Falls between Canada and the United States is diverted to a pumped storage reservoir during the night and released through hydroelectric generators during the day when the energy demand, and hence price, is higher.

In addition to single reservoirs, reservoir systems also exist. These may be (1) cascade reservoirs, consisting of a series of reservoirs constructed along a single river; or (2) inter-basin transfer schemes, designed to move water through a series of reservoirs, tunnels and/or canals from one drainage basin to another. These types are illustrated in Figure A.24.

Much of our current limnological knowledge (including that used to manage lakes and reservoirs) has come from studies of lakes over many decades. Although we now have a reasonable understanding of physical and biological processes that take place in lakes, we are less advanced in our understanding of these processes in reservoirs. Many early reservoir studies focused on sediment loading from drainage basins. The rationale for this was that the rate at which a reservoir filled with sediment is a major determinant of useful operational life. Comparatively little attention was given to the environmental and socioeconomic issues associated with reservoir construction. That situation has changed.

### 3.3.1. Shape and Morphometry

The shape of lakes and reservoirs is determined largely by how they were formed. This also affects some of their fundamental characteristics. Because lakes are naturally-formed, bowl-shaped depressions typically located in the

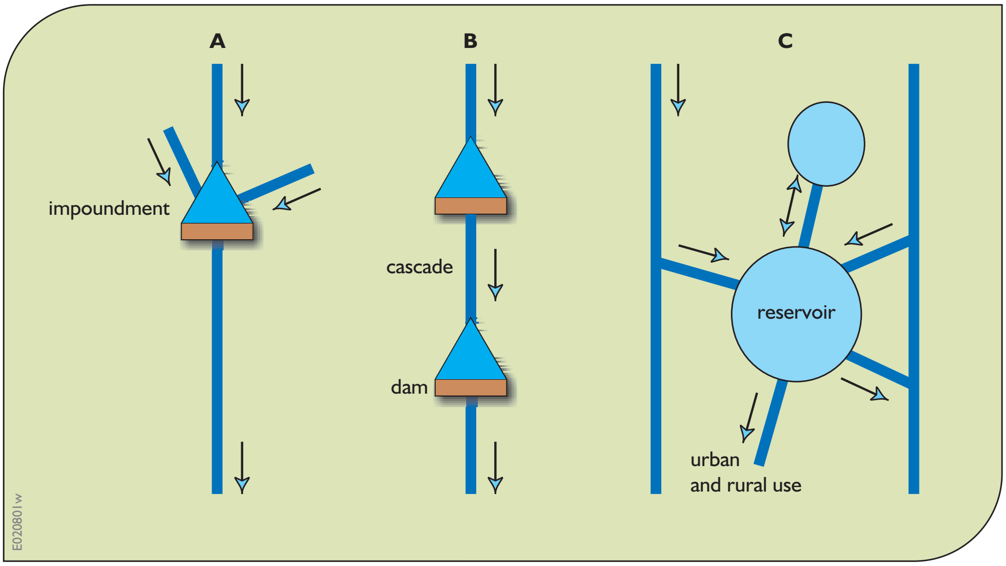



**Figure A.24.** Types of reservoir arrangements, ranging from (left to right) a single reservoir to a cascade of reservoirs along a river, to pumped storage reservoirs adjacent to rivers.

central part of a drainage basin, they usually have a more rounded shape than reservoirs. As in a bowl, the deepest part of a lake is usually at its centre. The shallowest part of the water basin is usually located near the outflow channel.

In contrast, a reservoir often has its deepest part near the dam. Moreover, because a river often has a number of streams or tributaries draining into it, when it is dammed the impounded water tends to back up into the tributaries. As a result, some reservoirs have a characteristic dendritic shape, with the 'arms' radiating outward from the main body of the reservoir as illustrated in Figure A.25. In contrast, a reservoir formed by damming a river with high banks will tend to be long and narrow. Depending on how they were constructed, off-river storage reservoirs can have many shapes.

The dendritic or branching form of many reservoirs provides a much longer shoreline than associated with lakes of similar volume. Reservoirs also usually have larger drainage basins. Because of their larger basins and multiple tributary inputs, the flow of water into reservoirs is more directly tied to precipitation events in the drainage basin than it is in lakes. Also, the fact that the deepest parts of most reservoirs are just upstream of the dam facilitates the possibilities for draining the reservoir.

Damming a river inundates land previously above water, and sometimes forces the relocation of inhabitants and wildlife living around the river. The presence of a dam downstream also allows a greater degree of control of water levels and volumes for reservoirs than for lakes. Constructing water discharge structures at different levels in the dam allows withdrawal or discharge of water from selected depths in a reservoir. 'Selective withdrawal' (as shown in Figure A.26) has major implications for the water-mixing characteristics and increases flushing possibilities for reservoirs.


**Figure A.25.** Characteristic dendritic shape of a large reservoir created by a dam to the right of this figure.

### 3.3.2. Water Quality

The characteristics of river water typically undergo changes as the water enters the lake or reservoir. Primarily because of reduced velocities, sediment and other

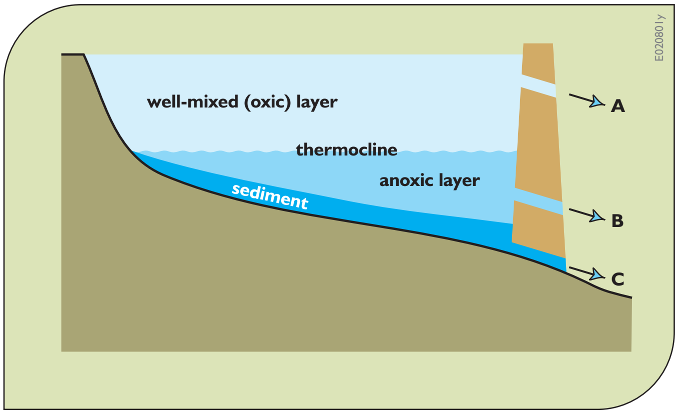



**Figure A.26.** Multiple outlets from a dam permit some control of water quality as well as quantity downstream.

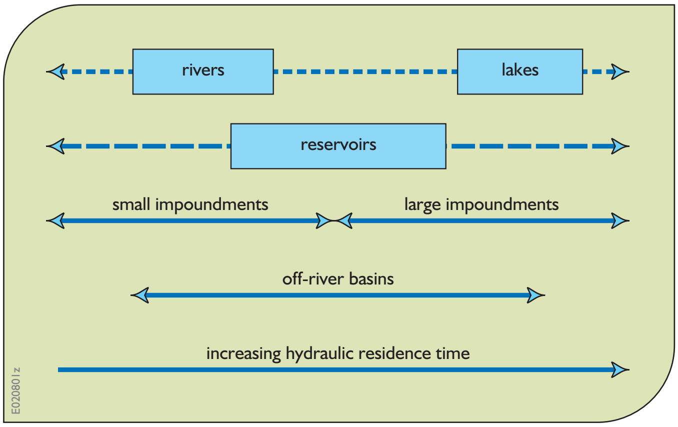



**Figure A.27.** A comparison of typical hydraulic residence times of rivers, reservoirs and lakes.

materials carried in the water settle out in the lake or reservoir. The structure of the biological communities also changes from organisms suited to living in flowing waters to those that thrive in standing or pooled waters. Pooled waters provide greater opportunities for the growth of algae (phytoplankton) that can lead to eutrophication.

Reservoirs typically receive larger inputs of water, as well as soil and other materials carried in rivers, than do lakes. As a result, they usually receive larger pollutant loads. However, because of greater water inflows, flushing rates are typically more rapid than in lakes (Figure A.27). Thus, although reservoirs may receive greater pollutant loads, they have the potential to flush the pollutants more rapidly. Reservoirs may therefore exhibit fewer or less severe negative water quality or biological impacts than lakes for the same pollutant load.

Although there are many variables of limnological significance, water quality is typically characterized on the basis of such variables as water clarity or transparency, concentration of nutrients and algae, oxygen concentration, concentration of dissolved minerals and acidity.

Waste chemical compounds from industry, some with toxic or deleterious effects on humans and/or water-dependent products, can be part of the pollution load discharged into lakes and reservoirs. These loads can kill aquatic organisms and damage irrigated crops. The quantity of bacteria, viruses and other organisms in waste-receiving waters are a primary cause of water-borne disease. Although such organisms present a risk to human health worldwide, such risks are particularly severe in developing countries.

There are some major differences between deep and shallow lakes and reservoirs. Deep water bodies, particularly in non-tropical regions, usually have better water quality in their lower layers. Shallow water bodies do not exhibit this depth differentiation in quality. Their more shallow, shoreline areas have relatively poorer water quality, because those areas are where pollutant inputs are discharged and where there is a greater potential for disturbance of bottom muds. Thus, the water quality of a natural lake usually improves from the shoreline to the deeper, central part. In contrast, the deepest end of a reservoir is immediately upstream of the dam, so that water quality usually improves along the length of a reservoir, from the shallow inflow end to the deeper, 'lake-like' end near the dam.

Reservoirs, particularly the deeper ones, are also distinguished from lakes by the presence of a longitudinal gradient in physical, chemical and biological water quality characteristics from the upstream river-end to the downstream dam-end. Thus, reservoirs have been characterized as having three major zones: an upstream riverine zone, a downstream lake-like zone at the dam-end, and a transitional zone separating these two zones (Figure A.28). The relative size and volume of the three zones can vary greatly in a given reservoir.

### 3.3.3. Downstream Characteristics

The construction of a dam can change the downstream river channel below it. Because reservoirs act as sediment and nutrient traps, the water at the dam-end of a reservoir is typically of higher quality than the water entering it.

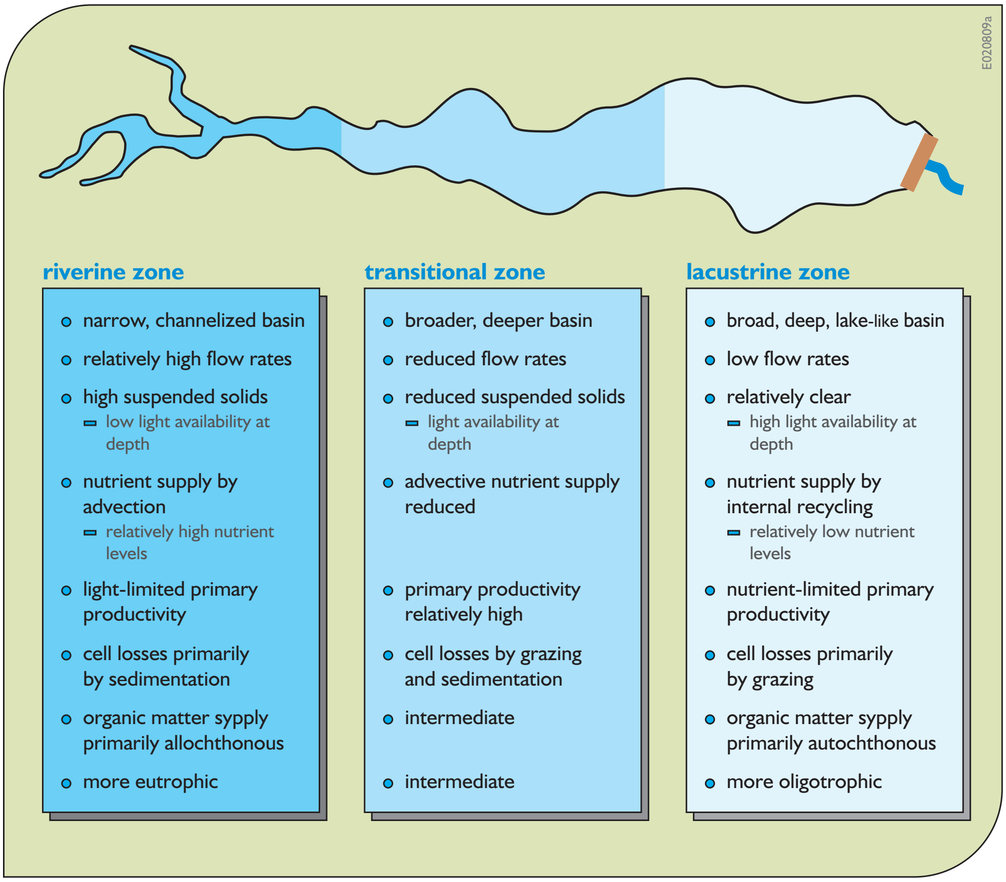



**Figure A.28.** Characteristics of three longitudinal zones in a reservoir.

This higher-quality water subsequently flows into the downstream river channel below the dam. The smaller the quantity of sediments and other materials transported in the discharged water, the greater the quantity of sediment that can be picked up and transported as it moves downstream. This scouring effect can have a negative impact on the flora, fauna and biological community structure in the downstream river channel. The removal and deposition of sediments from floodplains can also affect the biological character of those floodplains.

Many reservoirs, especially those used for drinking water supplies, have water release or discharge structures located at different vertical levels in their dams (Figure A.26). This allows for the 'selective withdrawal' or discharge of water from different layers within the reservoir. Depending on the quality of the water discharged, selective withdrawal can affect water quality within the reservoir itself, as well as the chemical composition and temperature of the downstream river. The ability to regulate or schedule water, temperature and silt discharges can provide some control on the hydrological regimes, affecting both flora and fauna.

Constructing a reservoir to protect downstream areas from floods often has significant social and economic implications, including the potential for stimulating urban and agricultural development adjacent to, and below, the reservoir. This can have both positive and negative impacts, depending on the nature and size of development. Appendix D discusses this aspect of flood management in more detail.

Table A.2 summarizes some of the major characteristics of lakes and reservoirs.

## 3.4. Management of Lakes and Reservoirs

Because drainage basins are simultaneously the sources of water, the places where the water is usually used, and the points where human activities affect both water quantity and quality, they are usually considered the logical management unit for lakes and reservoirs. However, activities that generate pollutants (such as urbanization, industrialization and agricultural production) may occur outside the affected drainage basin. They still need to be considered when developing management plans for lakes or reservoirs.

| lake | reservoir |
|---|---|
| especially abundant in glaciated areas; orogenic areas are characterized by deep, ancient lakes; riverine and coastal plains are characterized by shallow lakes and lagoons | located worldwide in most landscapes, including tropical forests, tundra and arid plains; often abundant in areas with a scarcity of natural lakes |
| generally circular water basin | elongated and dendritic water basin |
| drainage: surface area ratio usually $<10:1$ | drainage: surface area ratio usually $>10:1$ |
| stable shoreline (except for shallow, lakes in semi-arid zones) | shoreline can change because of ability to artificially regulate water level |
| water level fluctuation generally small (except for shallow lakes in semi-arid zones) | water level fluctuation can be great |
| long water flushing time in deeper lakes | water flushing time often short for their depth |
| rate of sediment deposition in water basin is usually slow under natural conditions | rate of sediment deposition often rapid |
| variable nutrient loading | usually large nutrient loading their depth |
| slow ecosystem succession | ecosystem succession often rapid |
| stable flora and fauna (often includes endemic species under undisturbed conditions) | variable flora and fauna |
| water outlet is at surface | water outlet is variable, but often at some depth in water column |
| water inflow typically from multiple, small tributaries | water inflow typically from one or more large rivers |

E020903h

**Table A.2.** Comparison of major characteristics of lakes and reservoirs.

Point sources of pollutants are 'pipeline' discharges to receiving waters. These are relatively easy to identify and isolate. In contrast, non-point pollution results from storm runoff or snowmelt that transports polluting materials diffusely and over urban or agricultural lands to streams, rivers, lakes and reservoirs. Non-point source pollution is closely tied to precipitation and runoff events. It is less predictable and more variable than pollutants from point sources and groundwater flows. Because of their diffuse nature, pollutants from non-point sources are more difficult to identify and control. Chapter 13 contains a more detailed discussion of this problem, especially from urban areas.

Of particular importance in addressing lake and reservoir management problems is the need to consider the affected ecosystems as well as the direct economic benefits lakes and reservoirs can provide. Managers must balance the water needs of economic development with the need to protect and preserve the environment and its ecosystem.

Many environmental and ecosystem processes in reservoirs and downstream rivers are complex, long term and 'non-traditional' from the perspective of current understanding of lake and river limnology. Some non-traditional processes exist because reservoirs are often intermediate aquatic systems representing transitions between flowing rivers and lakes (Figure A.28). Our limnological understanding of these processes is relatively young.

The presence of a reservoir in a drainage basin where no such water body previously existed can obviously affect the watercourse, its flora and fauna, and the human inhabitants in the drainage basin. These potential impacts should be identified and analysed prior to reservoir construction. The results of procedures to identify and properly evaluate potential environmental, social and economic consequences of reservoir construction are included in environmental impact assessment reports. In many countries such assessments are now required by law for all new dam construction.

### 3.5. Future Reservoir Development

Nearly all major river systems in the world have reservoirs in their drainage basins. A number of river systems (such as the Angara, Columbia, Dnieper, Missouri, Mosel, Parana and Volga) also have cascades of reservoirs within their basins. Reservoirs exist on all continents (except Antarctica) and in all countries, although their distribution within specific countries and regions is irregular. Construction of new reservoirs has all but ceased in North America and Europe. In contrast, reservoir construction is continuing in developing countries, with nearly all new reservoirs scheduled to enter operation in the twenty-first Century located in the Middle East, Asia, Africa and Latin America.

Reservoirs represent important components of the social and economic infrastructure of both developed and developing countries. In some cases, these have generated public and international concern, as in the cases of the Sardar Sarovar Dam in India and the Three Gorges Dam in China. Proponents of large dams argue that they bolster local economies, improve electrical energy supplies, provide needed flood control, and help humans manage the world's water resources more effectively. Opponents of large dams say that they cause significant damage to the environment and the local culture, and produce little overall economic gain. These conflicting points of view require attention early in the planning stage to ensure that they are properly considered by all relevant parties and interests prior to initiation of construction activities.

The Sanmenxia Dam on the Yellow River, China, provides an example of problems that were not sufficiently considered prior to dam construction. Finished in 1960, the project's goals were to prevent floods, provide water for irrigation and produce hydroelectric power. However, significant silt loads in the Yellow River were not adequately considered in the planning stage. The reservoir water basin was largely filled with silt only four years after construction, and the reservoir was subsequently taken out of operation.

Another example is the construction of the Aswan High Dam that impounds the Nile River in southern Egypt. The dam has now been in operation for about half a century, and both positive and negative impacts have resulted from its construction. The positive economic impacts include an improvement of summer crop rotations and guaranteed availability of irrigation water for agricultural production, expanded rice cultivation, conversion of about a million acres from seasonal to perennial irrigation, an expansion of about 486 thousand hectares of new agricultural and industrial land due to increased water availability, protection from high floods and droughts, generation of significant quantities of hydroelectric power, improved navigation possibilities and increased tourism.

The negative environmental and social impacts include declining water levels at Nile River barrages downstream of the dam, rising water levels upstream of the Delta Barrage, increased riverbank erosion and river meandering, production of river channel scour holes downstream of existing river barrages, decreased water quality due to increased industrial and agricultural discharges, increased reservoir siltation, increased reservoir eutrophication, increased water evaporation, increased coastal erosion at the mouth of the Nile River, decreased human health due to increased incidence of schistosomiasis and spread of water-related vectors, and inundation of historical monuments.

The overall conclusion of most observers is that the Aswan High Dam has had an overall positive effect, although it contributed to some significant environmental problems as well. Continued studies will undoubtedly provide additional information and guidance to those considering the construction of large dam projects in future years, particularly in developing countries.

Table A.3 summarizes some of the possible effects, both beneficial and damaging, of large reservoirs.

## 4. Wetlands

Wetlands are areas of frequent and prolonged presence of water at or near the soil surface. Swamps, marshes and bogs are well-recognized types of wetlands. Others less known include vernal pools (pools that form in the spring rains but are dry at other times of the year), playas (areas at the bottom of undrained desert basins that are sometimes covered with water) and prairie potholes. Some well-known wetlands, such as the Everglades in South Florida and the Mississippi bottomland hardwood swamps in the United States, can be completely dry at times. In contrast, many upland areas that are not considered wetlands can be very wet during and shortly after wet weather.

| positive benefits | negative effects |
|---|---|
| • production of energy (hydropower)<br>• increased low-energy water quality improvement<br>• retention of water resources in the drainage basin<br>• creation of drinking water and water supply resources<br>• creation of representative biological diversity reserves<br>• increased welfare for local population<br>• enhanced recreational possibilities<br>• increased protection of downstream river from flooding events<br>• increased fishery possibilities<br>• storage of water for use during low-flow periods<br>• enhancement of navigation possibilities<br>• increased potential for sustained agricultural irrigation | • displacement of local populations following inundation of reservoir water basin<br>• excessive human immigration into reservoir region, with associated social, economic and health problems<br>• deterioration of conditions for original population<br>• increased health problems from increasing spread of waterborne disease and vectors<br>• loss of edible native river fish species<br>• loss of agricultural and timber lands<br>• loss of wetlands and land/water ecotones<br>• loss of natural floodplains and wildlife habitats<br>• loss of biodiversity, and displaced wildlife populations<br>• need for compensation for loss of agricultural lands, fishery grounds and housing<br>• degradation of local water quality<br>• decreased river flow rates below reservoir, and increased flow variability<br>• decreased downstream temperatures, transport of silt and nutrients<br>• decreased concentrations of dissolved oxygen and increased concentrations of hydrogen sulfide and carbon dioxide in reservoir bottom water layer and dam discharges<br>• barrier to upstream fish migration<br>• loss of valuable historic or cultural resources (e.g., burial grounds, relic sites, temples)<br>• decreased aesthetic values<br>• increased seismic activity |

E020903j

**Table A.3.** Possible positive and negative effects of large reservoir construction.

## 4.1. Characteristics of Wetlands

When the upper part of the soil is saturated with water at growing season temperatures, organisms consume the oxygen in the soil and cause conditions unsuitable for most plants. Such conditions also cause the development of so-called ‘hydric soils’. Plants that can grow in these conditions, such as marsh grasses, are called ‘hydrophytes’. The presence of hydric soils and hydrophytes can indicate the existence of wetlands.

Wetlands are classified as either marine, estuarine, lacustrine, riverine or palustrine. Marine and estuarine wetlands are associated with the ocean and include coastal wetlands, such as tidal marshes. Lacustrine wetlands are associated with lakes, while riverine wetlands are found along rivers and streams. Palustrine wetlands may be isolated or connected wet areas and include marshes, swamps and bogs (Cowardin et al., 1979). A palustrine system can exist directly adjacent to or within lacustrine, riverine or estuarine systems.

Wetlands store precipitation and surface water and then slowly release the water into associated surface water resources, groundwater and the atmosphere. Wetland types differ in this capacity based on a number of physical and biological characteristics, including landscape position, soil saturation, the fibre content/degree of

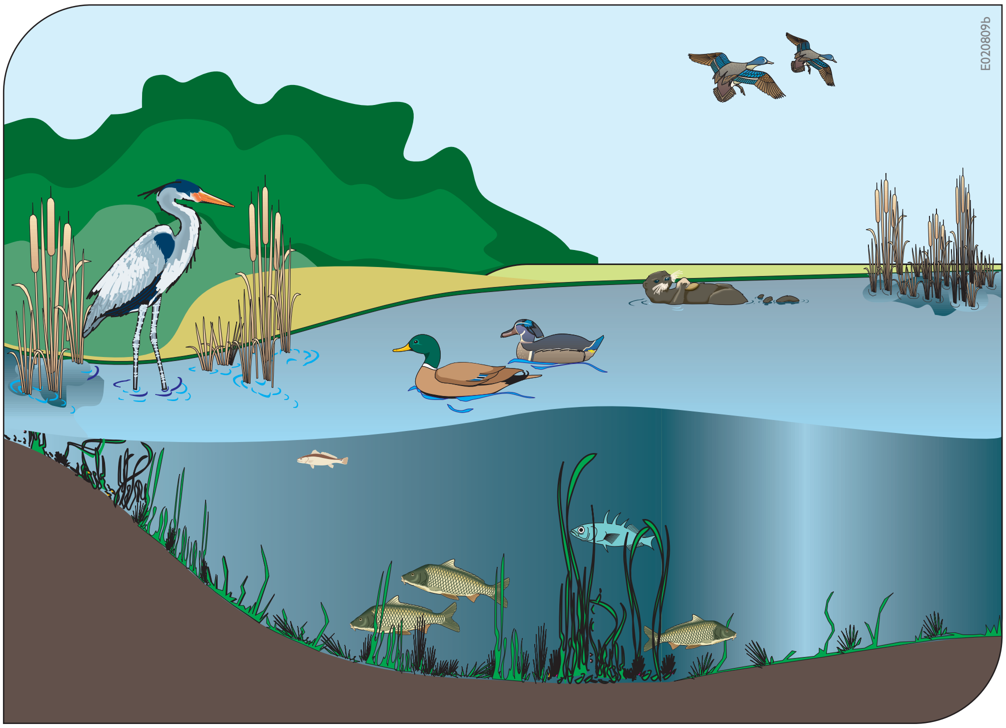



**Figure A.29.** Wetlands support a productive food web, from microscopic algae and submerged vascular plants to, in some areas, great blue herons and otters.

decomposition of the organic soils, vegetation density and type of vegetation.

### 4.1.1. Landscape Position

Landscape position affects the amount and source of water in a wetland. For example, wetlands that are near the top of a watershed, such as a mountain bog, will not receive as much runoff as marshes in low areas. Wetlands can result from precipitation, groundwater or surface flow. Precipitation-dominated wetlands can be in flat or slightly elevated areas in the landscape, where they receive little or no surface runoff. Generally such wetlands have a clay and peat layer that retains the precipitation and also prevents discharge from groundwater. Wetlands also form in areas of active groundwater discharge, particularly at the base of hills and in valleys. These groundwater-dominated wetlands may also receive overland flow but they have a steady supply of water from and to groundwater. Many wetlands in low points on the landscape are dominated by overland flow. Such riverine, fringe (marsh) and tidal wetlands actively influence the landscape since they come in contact with, store or release large quantities of water and act upon sediments and nutrients. Surface water provides the major source of water for these wetlands.

### 4.1.2. Soil Saturation and Fibre Content

Soil saturation and fibre content affect the capacity of a wetland to retain water. As in a sponge, as the pore spaces in wetland soil and peat become saturated by water, they are able to hold less additional water and release the water more easily. Clay soils retain more water than loam or sand. They hold the water particles more tightly through capillary action, since pore spaces are small and the water particles are attracted to the negatively-charged clay. Pore spaces between sand particles are larger and water drains more freely, since less of the water in the pores is close enough to be attracted to the soil particle.

Water in wetlands flows over or close to the surface in the fibric layer and root zone. Wetlands with sapric peat (mostly decomposed, unrecognizable fibres) and clay substrate will store water but will have no groundwater inflow or outflow.

### 4.1.3. Vegetation Density and Type

Plants growing in wetlands transpire water. Their stems reduce the velocity of water flowing through the wetland. As vegetation density increases, flow velocity decreases. Sturdy plants, such as shrubs and trees, cause more friction than do grasses.

Plant transpiration reduces the amount of water in wetland soil and increases the capacity to absorb additional water. As a result, water levels and outflows from the wetland are less during the growing season than when plants are dormant. Obviously larger plants and plants with greater surface area will take up and transpire more water than will smaller or less dense plant communities.

#### 4.1.4. Interaction with Groundwater

Some wetlands in low-lying areas are fed by groundwater discharges. Most other wetlands are a source of groundwater recharge. The extent of groundwater recharge by a wetland is dependent upon soil, vegetation, site, perimeter–volume ratio and water table gradient. Groundwater recharge occurs through mineral soils found primarily around the edges of wetlands; the soil under most wetlands is relatively impermeable. A high perimeter–volume ratio, as found in small wetlands, means that the surface area through which water can infiltrate into the groundwater is relatively high. Groundwater recharge is typical in small wetlands such as prairie potholes.

#### 4.1.5. Oxidation–Reduction

The fluctuating water levels (also known as hydrological flux) that are characteristic of wetlands control the oxidation–reduction (redox) processes that occur. These redox processes, governed by the frequency, duration and timing of wetland flooding (hydroperiod), play a key role in nutrient cycling, availability and export; pH; vegetation composition; sediment and organic matter accumulation; decomposition and export; and metal availability and export.

When wetland soil is dry, microbial and chemical processes occur, using oxygen as the electron acceptor. When wetland soil is saturated with water, microbial respiration and biological and chemical reactions consume available oxygen. The soil changes from an aerobic to an anaerobic, or reduced, condition. As conditions become increasingly reduced, other electron acceptors than oxygen must be used for reactions. These acceptors are, in order of microbial preference, nitrate, ferric iron, manganese, sulphate and organic compounds.

Wetland plants are adapted to changing redox conditions, and often contain spongy tissue with large pores in their stems and roots. This allows air to move quickly between the leaf surface and the roots, from which oxygen is released to oxidize the root zone (rhizosphere) and allow processes requiring oxygen, such as organic compound breakdown, decomposition and denitrification, to occur.

#### 4.1.6. Hydrological Flux and Life Support

Changes in the frequency, duration and timing of hydroperiods may affect spawning, migration, species composition and food chain support of the wetland and associated downstream systems. Normal hydrological flux allows exchange of nutrients and detritus and passage of aquatic life between systems.

Values of, or services provided by, wetlands as a result of the functions of hydrological flux and storage include enhanced water quality, water supply reliability, flood control, erosion control, wildlife support, recreation, and cultural and commercial benefits.

### 4.2. Biogeochemical Cycling and Storage

Wetlands may serve as sinks for nutrients, organic compounds, metals and components of organic matter. They may also act as filters of sediments and organic matter. A wetland may be a permanent sink for these substances if the compounds become buried in the substrate or are released into the atmosphere. Alternatively, it may retain them only during the growing season or under flooded conditions. Wetland processes play a role in the cycles of carbon, nitrogen and sulphur by transforming them and releasing them into the atmosphere (see, for example, the nitrogen cycle in Figure A.21).

Decomposition rates vary across wetland types, particularly as a function of climate, vegetation types, available carbon and nitrogen, and pH. A pH above 5.0 is necessary for bacterial growth and survival. Liming, to increase pH, accelerates decomposition, causing the release of carbon dioxide from wetlands and land subsidence.

The nutrients and compounds released from decomposing organic matter may be exported from the wetland in soluble or particulate form, incorporated into the soil, or eventually transformed and released to the atmosphere. Decomposed matter (detritus) forms the base of the aquatic and terrestrial food web.

Decomposition requires oxygen and thus reduces the dissolved oxygen content of the water. High rates of decomposition – such as after algae blooms – can reduce water quality and impair aquatic life support.

The values of wetland functions related to biogeochemical cycling and storage include water quality and erosion control.

### 4.2.1. Nitrogen (N)

The biological and chemical process of nitrification and denitrification in the nitrogen cycle transforms the majority of organic nitrogen entering wetlands, causing between 70% and 90% to be removed.

In aerobic substrates, organic nitrogen may mineralize to ammonium, which plants and microbes can use, adsorb to negatively charged particles (e.g. clay), or diffuse to the surface. As ammonia ($NH_3$) diffuses to the surface, the bacteria *Nitrosomonas* can oxidize it to nitrite ($NO_2$). The bacteria *Nitrobacter* in turn oxidizes nitrite to nitrate ($NO_3$). This process is called nitrification. Conversely, plants or microorganisms can assimilate nitrate, or anaerobic bacteria may reduce it (denitrification) to gaseous nitrogen ($N_2$) when nitrate diffuses into anoxic (oxygen depleted) water. The gaseous nitrogen volatilizes and the nitrogen is eliminated as a water pollutant. Thus, the alternating reduced and oxidized conditions of wetlands complete the needs of the nitrogen cycle and increase denitrification rates.

### 4.2.2. Phosphorus (P)

Phosphorus can enter wetlands attached to sediment or in dissolved form. Its removal from water in wetlands occurs through uptake of phosphorus by plants and soil microbes; adsorption by aluminium and iron oxides and hydroxides; precipitation of aluminium, iron, and calcium phosphates; and burial of phosphorus adsorbed to sediments or organic matter. Wetland soils can, however, reach a state of phosphorus saturation, after which phosphorus may be released from the system. Phosphorus is released into surface water as organic matter decomposes.

Dissolved phosphorus is processed by wetland soil microorganisms, plants and geochemical mechanisms. Microbial removal of phosphorus from wetland soil or water is rapid and highly efficient; following cell death, however, the phosphorus is released into the water again. Similarly, for plants, litter decomposition causes a release of phosphorus. Burial of litter in peat can provide long-term removal and storage of phosphorus. Harvesting of plant biomass is needed to maximize biotic phosphorus removal from the wetland system.

### 4.2.3. Carbon (C)

Wetlands store carbon within peat and soil, and so play an important role within the carbon cycle, particularly given observations of increasing levels of carbon dioxide in the atmosphere and concerns about global warming. When wetlands are drained the oxidizing conditions increase organic matter decomposition, thus increasing the release of carbon dioxide. When wetlands are preserved or restored, the wetlands act as a sink for carbon since organic matter decomposition is stabilized or slowed.

### 4.2.4. Sulphur (S)

Wetlands are capable of reducing sulphate to sulphide. Sulphide is released to the atmosphere as hydrogen, methyl and dimethyl sulphides or is bound in insoluble complexes with phosphate and metal ions in wetland sediments. Dimethyl sulphide released from wetlands may act as a seed for cloud formation. Sulphate may exist in soils or enter wetlands through tidal flow or atmospheric deposition.

### 4.2.5. Suspended Solids

Wetlands filter suspended solids from water that comes into contact with wetland vegetation. Stems and leaves cause friction that affects the flow of the water, thus allowing settling of suspended solids and removal of related pollutants from the water column. Wetlands may retain sediment in the peat or permanently as substrate. Sediment deposition varies across individual wetlands and wetland types, as deposition depends upon the rate and type of water flow (channelized or sheet flow), particulate size and vegetated area of the wetland.

### 4.2.6. Metals

All soils contain at least a low concentration of metals, but in some locations human activities have resulted in metal levels high enough to cause health or ecological risks. Metals may exist in wetland soils or enter wetlands through surface or groundwater flow.

Wetlands can remove metals from surface and groundwater as a result of the presence of clays, humic materials (peats), aluminium, iron and/or calcium. Metals entering wetlands bind to the negatively ionized surface of clay particles, precipitate as inorganic compounds (including metal oxides, hydroxides and carbonates controlled by system pH), interact with humic materials, and adsorb or occlude to precipitated hydrous oxides. Iron hydroxides are particularly important in retaining metals in salt marshes. Wetlands remove more metals from slow-flowing water since there is more time for chemical processes to occur before the water moves out of the area. Burial in the wetland substrate will keep bound metals immobilized. Neutral pH favours metal immobilization in wetlands. With the exception of very low pH peat bogs, as oxidized wetland soils are flooded and reduced, pH converges toward neutrality (6.5 to 7.5), whether the wetland soils were originally acidic or alkaline.

## 4.3. Wetland Ecology

Wetlands are productive ecosystems. Immense varieties of species of microbes, plants, insects, amphibians, reptiles, birds, fish and other wildlife depend on them in some way. Wetlands with seasonal hydrological pulsing are the most productive.

Wetland plants play an integral role in the ecology of the watershed. They provide breeding and nursery sites, resting areas for migratory species, and refuge from predators. Decomposed plant matter (detritus) released into the water is the source of food for many invertebrates and fish both in the wetland and in associated aquatic systems. Physical and chemical characteristics such as climate, topography, geology, hydrology, and inputs of nutrients and sediments determine the rate of plant growth and reproduction (primary productivity) of wetlands.

The greater the amount of vegetation, the more the wetland vegetation will intercept runoff and be capable of reducing runoff velocity and removing pollutants from the water. Wetland plants also reduce erosion as their roots hold the soil particles of streambanks, shorelines or coastlines.

The inundated or saturated conditions occurring in wetlands limit plant species composition to those that can tolerate such conditions. Beaver, muskrat and alligators create or manipulate their own wetland habitat, which other organisms, such as fish, amphibians, waterfowl, insects and mammals, can then inhabit.

Wetland shape and size affect the habitat of the wildlife community. The shape affects the perimeter–area ratio, which is important for the success of interior and edge species. Shape is also important for movement of animals within the habitat and between habitats. Wetland size is particularly important for larger and wider-ranging animals that use wetlands for food and refuge, such as black bear or moose, since in many regions wetlands may be the only undeveloped and undisturbed areas remaining.

## 4.4. Wetland Functions

Only recently have scientists begun to understand the importance of the functions that wetlands perform. Far from being useless disease-ridden places, wetlands provide benefits that no other ecosystem can, including natural water quality improvement, flood protection, shoreline erosion control, opportunities for recreation and aesthetic appreciation, community structure and wildlife support, and natural products at reduced costs compared with other alternatives. Protecting wetlands in turn can increase human safety and welfare as well as enhance the productivity of aquatic ecosystems.

### 4.4.1. Water Quality and Hydrology

Wetlands have important filtering capabilities for intercepting surface water runoff from higher dry land before it reaches open water. As the runoff water passes through, the wetlands retain excess nutrients and some pollutants, and reduce sediment that would otherwise clog waterways and affect fish and amphibian egg development. In performing this filtering function, wetlands offset the costs of wastewater treatment. In addition to improving water quality through filtering, some wetlands maintain streamflow during dry periods, and many replenish groundwater supplies.

### 4.4.2. Flood Protection

Wetlands function as natural sponges. They trap and slowly release surface water, rain, snowmelt, groundwater and floodwaters. Trees, root mats and other wetland vegetation slow the speed of floodwaters and distribute them over the floodplain. This combined water storage and braking action lowers flood heights and reduces erosion. Wetlands within and downstream of urban areas are particularly valuable, counteracting the greatly increased rate and volume of surface water runoff from pavement and buildings.

The holding capacity of wetlands helps control floods and prevents waterlogging of crops. Preserving and restoring wetlands, together with other water retention measures, can often provide the level of flood control that would otherwise require expensive dredging operations and levees. The bottomland hardwood-riparian wetlands along the Mississippi River once stored significantly more floodwater than they can today because most have been filled or drained.

### 4.4.3. Shoreline Erosion

The ability of wetlands to control erosion is so valuable that communities in some coastal areas are restoring wetlands to buffer the storm surges from hurricanes and tropical storms. Wetlands at the margins of lakes, rivers, bays and the ocean protect shorelines and stream banks against erosion. Wetland plants hold the soil in place with their roots, absorb the energy of waves, and break up the flow of stream or river currents.

### 4.4.4. Fish and Wildlife Habitat

Many animals and plants depend on wetlands for survival. Estuarine and marine fish and shellfish, various birds and certain mammals depend on coastal wetlands for survival. Most commercial and game fish breed and raise their young in coastal marshes and estuaries. Menhaden, flounder, sea trout, spot, croaker and striped bass are among the more familiar fish that depend on coastal wetlands. Shrimp, oysters, clams, and blue and Dungeness crabs likewise need these areas for food, shelter and breeding grounds.

For many animals and plants, like wood ducks, muskrat, cattails and swamp rose, inland wetlands are the only places they can live. Beaver may actually create their own wetlands. For others, such as striped bass, peregrine falcon, otter, black bear, raccoon and deer, wetlands provide important food, water or shelter. Many breeding bird populations – including ducks, geese, woodpeckers, hawks, wading birds and many songbirds – feed, nest and raise their young in wetlands. Migratory waterfowl use coastal and inland wetlands as resting, feeding, breeding or nesting grounds for at least part of the year. Indeed, an international agreement to protect wetlands of international importance was developed because some species of migratory birds are completely dependent on certain wetlands and would become extinct if those wetlands were destroyed.

### 4.4.5. Natural Products

We use a wealth of natural products from wetlands, including fish and shellfish, blueberries, cranberries, timber and wild rice, as well as medicines that are derived from wetland soils and plants. Many fishing and shellfishing industries harvest wetland-dependent species. In the southeast part of the United States, for example, nearly all the commercial catch and over half of the recreational harvest consists of fish and shellfish that depend on the estuary-coastal wetland system. Wetlands are habitats for fur-pelted animals like muskrat, beaver and mink, as well as reptiles such as alligators.

### 4.4.6. Recreation and Aesthetics

Wetlands have recreational, historical, scientific and cultural value. More than half of all US adults hunt, fish, go birdwatching or photograph wildlife. Painters and writers continue to capture the beauty of wetlands on canvas and paper, or through cameras, and video and sound recorders. Others appreciate these wonderlands through hiking, boating and other recreational activities. Almost everyone likes being on or near the water; part of the enjoyment is the fascinating variety of lifeforms.

## 5. Estuaries

Estuaries are places where freshwater mixes with saltwater under the influence of tides, creating a unique environment supporting a rich and diverse ecosystem.

Estuaries are capable of gathering and holding an abundance of life-giving nutrients from the land and from the ocean. They are also important sites for human economic and recreational activities. An estuary can provide a laboratory for lessons in biology, geology, chemistry, physics and social issues.

Estuaries are partially enclosed bodies of water formed where freshwaters from rivers and streams flow into and mix with the saline seawater. They and the lands surrounding them are places of transition from land to sea, and from fresh to saltwater. Although influenced by the tides, they are generally protected from the full force of ocean waves, winds and storms by the reefs, barrier islands or fingers of land, mud or sand that define their seaward boundaries.

Estuarine geometry can vary substantially, and the physical characteristics of estuaries affect the distribution of their constituents. The shape and geomorphology of the estuary influence the flows, circulation and mixing of the waters. The flow pattern and the density structure in an estuary govern the advective transport and diffusion/dispersion characteristics that determine the fate and distribution of waterborne constituents.

The word *estuary* comes from the Latin 'aestus', meaning tide. Traditionally, the upstream limit of an estuary is defined in terms of the limit of penetration of saltwater. Salt concentrations move upstream under the influence of the ocean tide. A commonly used definition is that of Pritchard (1952), who defined an estuary as 'a semi-enclosed coastal body of water which has a free connection with the open sea and within which sea water is measurably diluted with fresh water derived from land drainage'.

A broader definition of an estuary would take into account the diversity and spatial variability of its fauna and flora. Hutchings and Collett (1977) define estuaries as the tidal portions of river mouths, bays and coastal lagoons, irrespective of whether they are dominated by hypersaline, marine or freshwater conditions. Included in this definition are inter-tidal wetlands, where water levels can vary in response to the tidal levels of the adjacent waterway, together with perched freshwater swamps, as well as coastal lagoons that are intermittently connected to the ocean.

Perched freshwater swamps, or swamps located above the regional water table, can occur in the lower catchment areas of an estuary. Although tides or ocean salinity does not affect them, they can form an important component of the estuarine habitat. For this reason, such swamps need to be included in estuarine management plans and policies.

## 5.1. Types of Estuaries

Every estuary is unique, yet all share certain features. Some are similar enough to be grouped or classified according to their shared characteristics. Others are typical of specific regions. All are greatly affected by geology, climate and many other factors, including the ever-growing human population.

Depending on their geological characteristics, estuaries can be classified as:

- *Coastal plain estuary* (*or drowned river valley*). These are commonly found along coastlines that have relatively wide coastal plains. They were created by the gradual rise in sea level following the last glacial period, some 10,000 years ago. Usually, a stretch of the freshwater river is subject to tidal oscillations. Typically, the estuary is funnel-shaped, widening gradually in the downstream direction, often with inter-tidal mud flats. Sometimes a large delta system has formed with many natural channels. Chesapeake Bay is the largest estuary of this type of in the United States.
- *Fjord-type estuary* (*formed by a glacier*). These estuaries generally have U-shaped cross sections with steep sides and are often quite deep (300 or 400 m). They are common in the Northern Hemisphere above 45 degrees latitude. In North America they are most spectacular along the coast of Alaska and British Columbia.
- *Bar-built estuary or lagoon*. These are formed when offshore barrier sand islands and sand pits become higher than sea level and extend between headlands in a chain, broken by one or more inlets. Lagoons are usually situated parallel to the coastline. Many have narrow outlets to the sea and minimal freshwater inflow, often creating higher salinity levels than in coastal-plain and other estuaries. In North America, lagoons occur mainly along the Gulf Coast. Smaller lagoons occur along the West Coast.
- *Estuaries produced by tectonic processes*. Estuaries created by processes such as landslides, faulting and volcanic eruptions are found along coasts where such

activity is or has been common, such as the Pacific Coast of North America. These estuaries vary greatly and often share characteristics with other types. The largest in the United States is San Francisco Bay.

Any existing estuary can be a 'composite estuary' evolving from an overlap of two or more of these basic types.

Estuaries can also be classified on the basis of stratification and circulation:

- A *salt-wedge* estuary is highly stratified. Saltwater moves into it in the shape of a wedge, with freshwater flowing over it. The velocity of the freshwater outflow is usually greater than that of the saltwater inflow. This keeps the saltwater from extending very far up the estuary. The Mississippi River Estuary is an example of this type, where the difference in water level between the high and low tides (the tidal range) on the seaward side of the estuary is small and incapable of mixing the stratified layers.
- The *vertically homogenous* or well-mixed estuary is characterized by low inflow of freshwater and large tidal ranges. In such estuaries, saltwater and freshwater tend to mix vertically, and sometimes laterally as well.
- *Intermediate* estuaries are partly mixed. They exhibit circulation patterns that are somewhere between those of the salt-wedge and vertically homogenous estuaries. This type of estuary, typical along the US East Coast, has a moderate inflow of freshwater and moderate-to-large tidal range.

Estuaries area also classified based on their circulation – the interaction of tidal currents and river flow (Bowden, 1967):

- A salt wedge estuary has only a small amount of friction between the layers. The steep density gradient at the interface reduces the turbulence and mixing to a low level. The effect of the Coriolis force causes the lateral sloping of the interface downward to the right in the Northern Hemisphere when looking towards the sea.
- A vertically homogeneous estuary can exhibit lateral variation due to Coriolis force in wide estuaries, or be laterally homogeneous where the ratio of width to depth is relatively small. Coriolis force causes a net seaward flow of lower-salinity water on the right hand of the estuary, and a compensating flow of higher-salinity water on the left-hand side. In this sequence, the tidal currents are stronger in relation to the river flow.
- In a two-layer flow with entrainment, the entrainment is due to breaking internal waves. Many fjords have this type of circulation.
- In a two-layer flow with vertical mixing, the volume of water involved in this type of circulation may be many times the river discharge. Examples include the Tees and Thames estuaries in the UK.

Estuaries are constantly affected by tidal and wave action, prevailing and changing winds, local and distant weather systems, and variations of rainfall runoff and river discharge.

## 5.2. Boundaries of an Estuary

With some exceptions, such as those found in the Mediterranean Sea, tidal motions are significant features in estuaries. The upstream boundary or head of an estuary is the limit of tidal influence. The tidal limit can be a considerable distance upstream from the salinity limit. The actual limit of tidal influence varies with time, depending upon freshwater flows and the natural variability of tides. It can be difficult to determine the tidal limit by mere observation. Towards the limit of tidal influence, flood flows may persist for as little as an hour or less, and may be so small as to go unnoticed.

As well as short-term cyclical changes in response to the changing ocean tides, there may be long-term variations in the upstream boundary of an estuary due to both natural processes and artificial disturbance. The downstream boundary is not always obvious, but it corresponds to the seaward limit of the entrance bar in most situations.

The lateral boundaries of an estuary are often defined in ecological rather than hydraulic terms. The shallow and inter-tidal margins of an estuary support the diversity and productivity of marine life, such as sea-grass meadows, mangrove forests and saltmarshes, all generating large amounts of organic detritus that form the foundation of the estuarine food chain.

The lateral extent of an estuary includes all wetlands – salt, brackish and fresh – that interact with tidal and flood flows. They also include those marshes that are inundated only during extreme tides or flood events.

## 5.3. Upstream Catchment Areas

An estuary acts as a funnel to convey freshwater runoff from the land into coastal waters. Catchment activities can affect the volume and quality of this runoff. Floodwaters and stormwater runoff often contain significant quantities of suspended solids, natural and artificial nutrients, pesticides and other constituents, some of which can be detrimental to estuarine ecosystems.

Upstream catchment activities are the single most important factor in determining the nutrient balance and water quality of estuaries. The impact of external activities occurring beyond the strictly defined estuarine limits points to the need for a *total catchment management* approach in managing estuaries. What we do upstream affects what happens downstream, including, for most rivers, the estuary.

## 5.4. Water Movement

Water moves along an estuary under the influence of two primary forcing mechanisms: freshwater inflows from rivers and the regular tidal movement of seawater into and out of the estuary. In addition, differences in tides, winds and salinity densities within the estuary generate secondary currents that, while of low velocity, can be significant with respect to mixing and sediment transport.

Freshwater inflows fluctuate in a more or less random fashion in response to surface runoff from tributary catchments. The construction of major dams in upstream catchment areas reduces both the volume of freshwater runoff and the freshwater flushing of estuaries. In times of drought, freshwater inflows may be absent, or may be mainly the discharge of sewage effluent and other wastewaters.

Freshwater inflows result in a net seaward excursion of water particles over each tidal cycle, and thus promote estuary flushing.

The movement of seawater in and out of an estuary is predominantly influenced by the tides. Freshwater effects are generally small (but often significant with respect to water quality), except during floods.

Coastal water levels fluctuate in a regular and predictable fashion in response to gravitational effects, primarily of the moon and sun, on the oceans of the earth. The tidal range varies from one tidal cycle to another in response to the changing relative positions of these celestial bodies. In response to the monthly orbit of the moon around the earth, the tidal range undergoes a regular fourteen-day cycle, increasing to a maximum over a week (spring tides) and then decreasing to a minimum over the following week (neap tides). Solstice tides, or king tides occur in June and December of each year, when the sun is directly over the Tropics of Cancer and Capricorn respectively.

Tides along many coastlines are semi-diurnal in nature; that is, high water and low water occur about twice daily (the actual period of a tidal cycle is about 12.5 hours). They are sinusoidal in shape and have a pronounced diurnal inequality (successive high tides differ markedly).

The tidal rise and fall of ocean water levels propagates along an estuary as a wave. The speed of travel or celerity of this wave – the speed at which high water and low water travel upstream from the estuary entrance – varies with water depth: the deeper the water, the faster the wave celerity. The propagation of the tide along an estuary is affected by the geometry of its bed. Tidal propagation in tidal rivers is very sensitive to water depths. In certain estuaries, the ocean tidal range is increased in upstream reaches of the estuary because of its geometry. The celerity of the flood tide is higher than that of the ebb tide.

### 5.4.1. Ebb and Flood Tides

The vertical rise and fall of tides (Figure A.30) produce horizontal flows in the form of tidal currents. The incoming or rising tide is traditionally referred to as the flood tide because it floods the channel. The outgoing tide is referred to as the ebb tide. The strength of the ebb and flood tide velocities varies diurnally and over spring–neap cycles in exactly the same way as tidal water levels do. Spring tides produce the fastest tidal currents. Depending on the tidal range, entrance characteristics and the depth of the estuary, tidal currents can have velocities of up to a metre a second.

If the tidal range is appreciable compared to the mean depth of the estuary, the speed of propagation of the tide at high water will be significantly faster than at

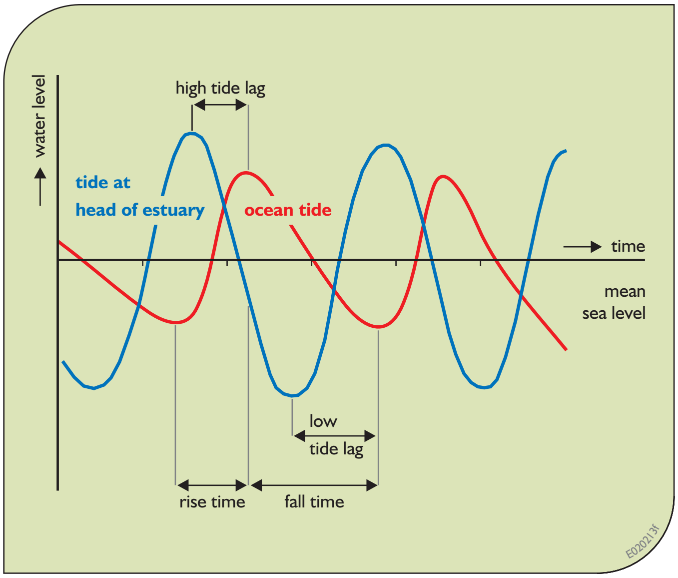



**Figure A.30.** Tidal characteristics at the entrance and head of a tidal river.

low water. This causes the shape of the tidal wave to become progressively more distorted as it moves landward. This tidal distortion results in a saw-tooth tide curve, in which the rise of the tide is noticeably faster than its fall. As a consequence, peak flood tide velocities are generally greater than peak ebb tide velocities. This is of significance with respect to sediment transport.

The period of quiet water when the tide reverses from flood to ebb or vice versa is referred to as slack water. High-water slack is the name given to the tide change from flood to ebb; low-water slack refers to the tide change from ebb to flood. The duration of slack water varies from one estuary to another and can last from twenty minutes to almost an hour.

### 5.4.2. Tidal Excursion

The total distance travelled by a water particle from low-water slack to high-water slack and vice versa is referred to as the tidal excursion. This is the maximum distance travelled by a water particle during the rising or falling limb of the tide.

Tidal excursion is not to be confused with the distance travelled by the tide wave itself (e.g. high water) that propagates from the ocean to the end of the estuary each tide cycle.

Freshwater inflows impose a net seaward movement on water particles over a tide cycle. In these circumstances, the ebb tide excursion is greater than the flood tide excursion. This is illustrated in Figure A.31, which also shows the net effect of oscillatory tidal flows and seaward-draining freshwater flows in flushing a parcel of water out to sea.

### 5.4.3. Tidal Prism

The total volume of water moving past a fixed cross section of the estuary during each flood tide or ebb tide (i.e. slack water to slack water) is referred to as the tidal prism. The larger the tidal range within the estuary and the greater the dimensions of the estuary, the larger is the tidal prism. On average, the ebb and flood tidal prisms are equal.

The size of the tidal prism is not an indication of the amount of flushing. The movement of water contained in the tidal prism is largely oscillatory, as depicted in Figure A.31. The net seaward flushing action over a tidal cycle results from two processes: freshwater advection and longitudinal dispersion. Freshwater advection is the net seaward displacement caused by Freshwater. Freshwater longitudinal dispersion results mainly from the tides.

### 5.4.4. Tidal Pumping

The higher celerity of the flood tide compared to the ebb tide results in a tendency for greater upstream movement of water on the flood tide compared to downstream movement on the ebb tide. This leads to a dynamic trapping of water in the upper reaches of the estuary. This effect is sometimes referred to as tidal pumping: tidal distortion results in the tide pumping water upstream.

In shallow estuaries, tidal pumping can affect the distribution of dissolved pollutants along an estuary.

### 5.4.5. Gravitational Circulation

The presence of salt in an estuary produces a longitudinal density gradient, with water densities around the mouth of the estuary being greater (because of higher salt concentrations) than densities around the head of the estuary. This results in the enhancement of flood-tide velocities near the bed and ebb-tide velocities near the surface. When averaged over a tidal cycle, this behaviour leads to

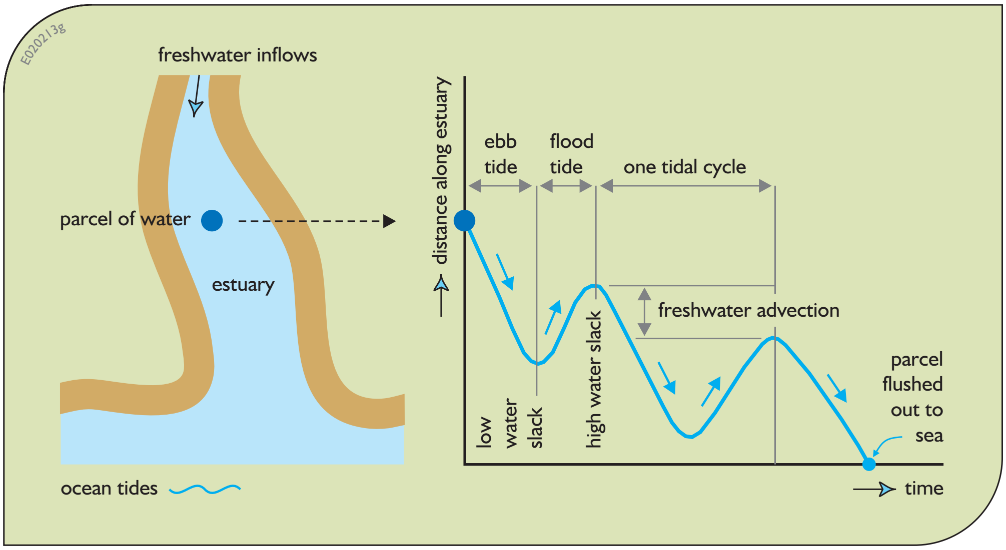



**Figure A.31.** Movement of a 'parcel' of water down an estuary.

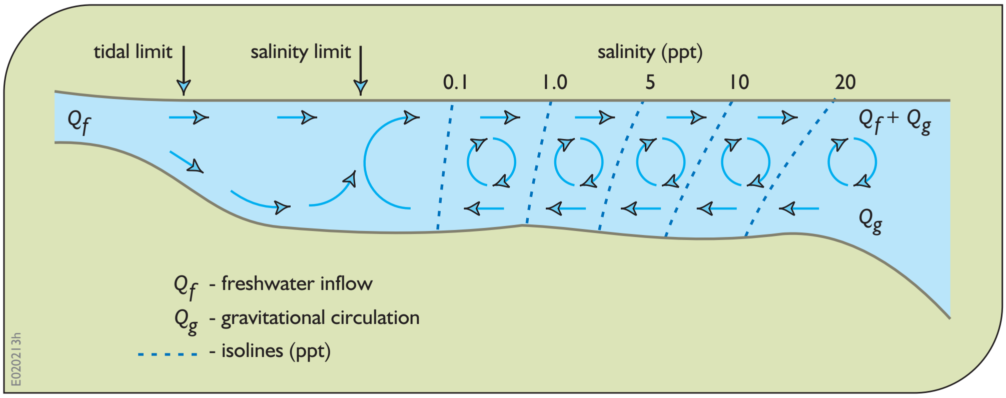



**Figure A.32.** Gravitational circulation in a partially mixed estuary.

residual currents, in which saline water flows upstream along the bottom of the estuary and less salty, even fresh water, flows seawards near the surface (see Figure A.32). This pattern of residual flows is referred to as gravitational circulation (it is driven by the gravitational forces resulting from density differences). Gravitation circulation can give rise to discharges considerably greater than freshwater inflows. Gravitational circulation is an important mechanism of upstream sediment transport and the longitudinal dispersion of salt in an estuary.

### 5.4.6. Wind-Driven Currents

Wind shear at the water surface produces surface currents, counterbalancing currents at the bottom of the water column, and large-scale lateral circulation. Wind speeds up to about 6 to 7 m/sec are the most effective in moving surface waters. Wind-driven surface currents have speeds of about 2% of the wind speed. Wind-driven currents are a major mechanism in the transport and distribution of floating pollutants such as oil.

In sluggish tidal systems or bays, where surface tidal flows are weak, wind-driven currents can be among the main agents leading to effective water movement and mixing within the main water body. When a constant wind blows over a basin of variable depth, a laterally varying surface current is induced, flowing with the wind in shallow areas and as a return flow against the wind in deeper areas.

## 5.5. Mixing Processes

Parcels of water mix as they move along the estuary under the influence of freshwater flows, tidal flows and secondary currents. Mixing not only involves an exchange of water mass, but also of any substance dissolved in it, such as salts and various pollutants. Mixing processes affect the distribution of salinity and water quality constituents throughout the estuary.

### 5.5.1. Advection and Dispersion

Advective and dispersive transport are the major processes by which dissolved matter is distributed throughout an estuary. As water flows along an estuary it transports dissolved matter with it. This is advective transport. During this process, water mixes with neighbouring parcels of water. This mixing, or dispersive transport, leads to a net transport of dissolved materials from regions of high concentration to regions of lower concentration.

The mixing of parcels of water is principally caused by lateral and vertical variations in velocity along an estuary. These variations create velocity shear. Velocity variations result in faster parcels of water – together with their dissolved loads – moving ahead of their slower neighbours, as depicted in Figure A.33. Turbulence and eddies generated by the velocity shear between neighbouring parcels result in the interchange of water and dissolved matter between parcels.

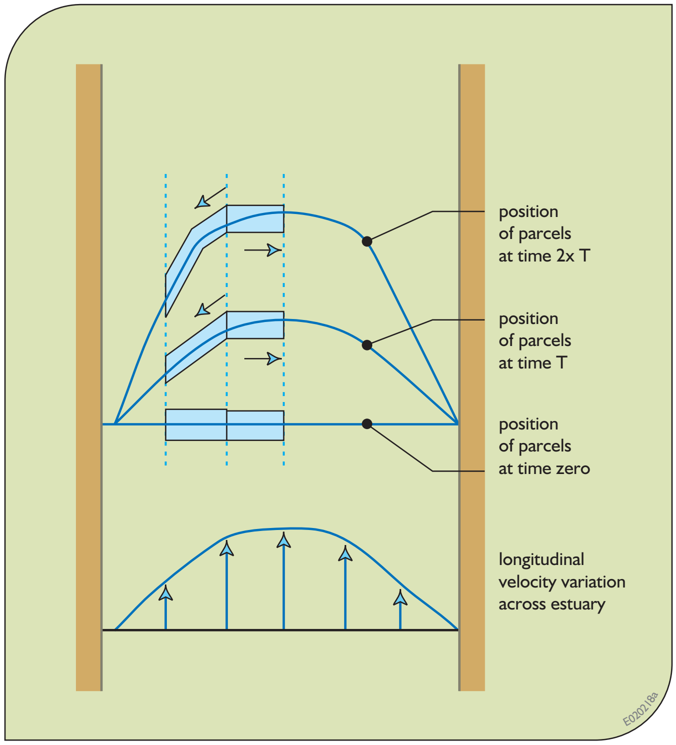



**Figure A.33.** Mixing caused by the lateral variation of velocity across an estuary.

### 5.5.2. Mixing

Mixing occurs in all three directions – vertically laterally and longitudinally. The principal mechanism of vertical mixing is the velocity shear caused by flow moving across the bed of the estuary. The surface and mid-depth flow velocities are faster than bed-flow velocities. This results in turbulence, principally at the bed, which promotes the mixing of overlying waters. Vertical mixing is responsible for well-mixed salinity regimes.

The major mechanisms promoting the mixing of waters laterally across an estuary are the velocity shear associated with lateral velocity gradients, wind shear, lateral tidal flows, large-scale eddies generated by obstructions, and bends and meanders in the alignment of major channels.

Figure A.33 shows the lateral exchange of parcels of water associated with the lateral variation of longitudinal velocities across the estuary. Lateral variations in velocity are caused by the presence of the banks of the estuary and changes in water depth. Deeper waters flow faster than shallow waters. This lateral exchange promotes mixing across the estuary.

Wind shear and lateral tidal flows (such as the filling of inter-tidal areas on the flood tide) both result in the advection of water across an estuary and associated mixing. The presence of bends and meanders also results in water flowing from one side of a channel to the other, and associated mixing across the channel.

Obstructions such as rock bars and shoals also promote the flow of water across an estuary, both by the advection associated with redirected flows and by the dispersion associated with lateral shear.

Large-scale boundary effects, tidal trapping, tidal loops and interconnections between tidal systems, and gravitational circulation are among a number of factors that promote longitudinal mixing. Large-scale boundary effects relate to the mixing caused by the presence of bays and channels along an estuary. Waters are temporarily stored in these areas on the rising tide and released when it falls. This separation of the trapped waters from other waters moving along the estuary facilitates mixing by velocity shear and advective transport.

The presence of shoals generates strong velocity shear between fast-moving main channel flows and the slow-moving waters that move in and out of shoal areas on the rising and falling tide. This behaviour leads to tidal trapping, whereby a discrete body of water is separated and trapped over shoal areas on the flood tide, to be released on the ebb. This behaviour facilitates mixing by velocity shear and advective transport, supplementing that resulting from large-scale boundary effects.

Finally, the presence of tidal loops in a large delta or the interconnection of two tidal systems can create a complicated pattern of residual flows that further enhances advective and dispersive transport along the estuary. To summarize, mixing will be greatest in those estuaries where lateral and vertical velocity gradients are greatest, that is, in estuaries where deep channels thread their way through shallow flats, where there are extensive shoals and peripheral bays and channels, and where tidal flows are fast.

## 5.6. Salinity Movement

Seawater consists of a dilute solution of a mixture of salts. The term salinity refers to the total salt concentration. Seawater has a worldwide average salt concentration of about 35 kg/m$^3$ or 35 parts per thousand (ppt). Saline coastal waters are carried into an estuary by the tides; freshwater inflows tend to wash the saltwater back out to sea. Tidal flows are very effective in moving salinity upstream. This is seen in the relatively quick migration of salinity upstream in an estuary after a major flood has washed it downstream. The balance between the landward transport of salt by tidal processes and its seaward return by freshwater discharges determines the limit of salinity intrusion along an estuary.

Salt and other dissolved substances are transported along an estuary by the longitudinal advection associated with large-scale water movements and secondary currents, and by the longitudinal dispersion associated with velocity shear. Freshwater inflow is the major factor affecting the limit of saline intrusion along an estuary. Salt will not penetrate far along channels with a net seaward residual velocity in excess of 0.1 m/sec.

### 5.6.1. Mixing of Salt- and Freshwaters

The density of seawater is greater than that of freshwater and varies depending on both salinity and temperature. At a temperature of 20 °C, seawater has a density of about 1,025 kg/m$^3$, whereas freshwater at that temperature has a density of 1,000 kg/m$^3$. Although the difference in density is relatively small, it still affects estuarine circulation.

Because of its lower density, freshwater tends to float on top of the seawater (stratification). Turbulence generated by the movement of the water over the bed of an estuary causes vertical mixing, which tends to break down any saline–freshwater stratification. The faster the water moves, the greater its turbulence and resultant mixing. High tidal velocities produce strong vertical mixing, resulting in little variation in salinity from the top to the bottom of the water column. Low tidal velocities are insufficient to cause complete vertical mixing and stratified conditions can develop (i.e. bottom salinity concentrations are greater than surface salinity concentrations).

### 5.6.2. Salinity Regimes

In well-mixed estuaries, the salinity distribution is almost uniform with depth. In partially mixed estuaries, the salinity varies continuously through the depth of the water column, with no evidence of a marked interface between the upper and lower layers. The salinity can vary over the depth by as little as 1 ppt or as much as 10 ppt.

Stratified conditions are characterized by an abrupt increase in salinity over the water depth. In the absence of significant tidal velocities, vertical mixing in an estuary is weak. This allows the saltwater to move into the estuary as an arrested salt wedge that can penetrate a long distance into a deep estuary.

The freshwater overlying the wedge tends to entrain saltwater as it moves seaward, thereby becoming more brackish as it approaches the estuary mouth. These conditions are shown in Figure A.34.

### 5.6.3. Variations due to Freshwater Flow

If there is considerable variation in freshwater inflows over time, the estuaries can display all of the foregoing salinity regimes (shown in Figure A.34) at different times. During periods of extended dry weather, when freshwater flows are small, estuaries tend to be well mixed. When freshwater inflows to estuaries are low, the salinity of water in the estuary may exceed the salinity of

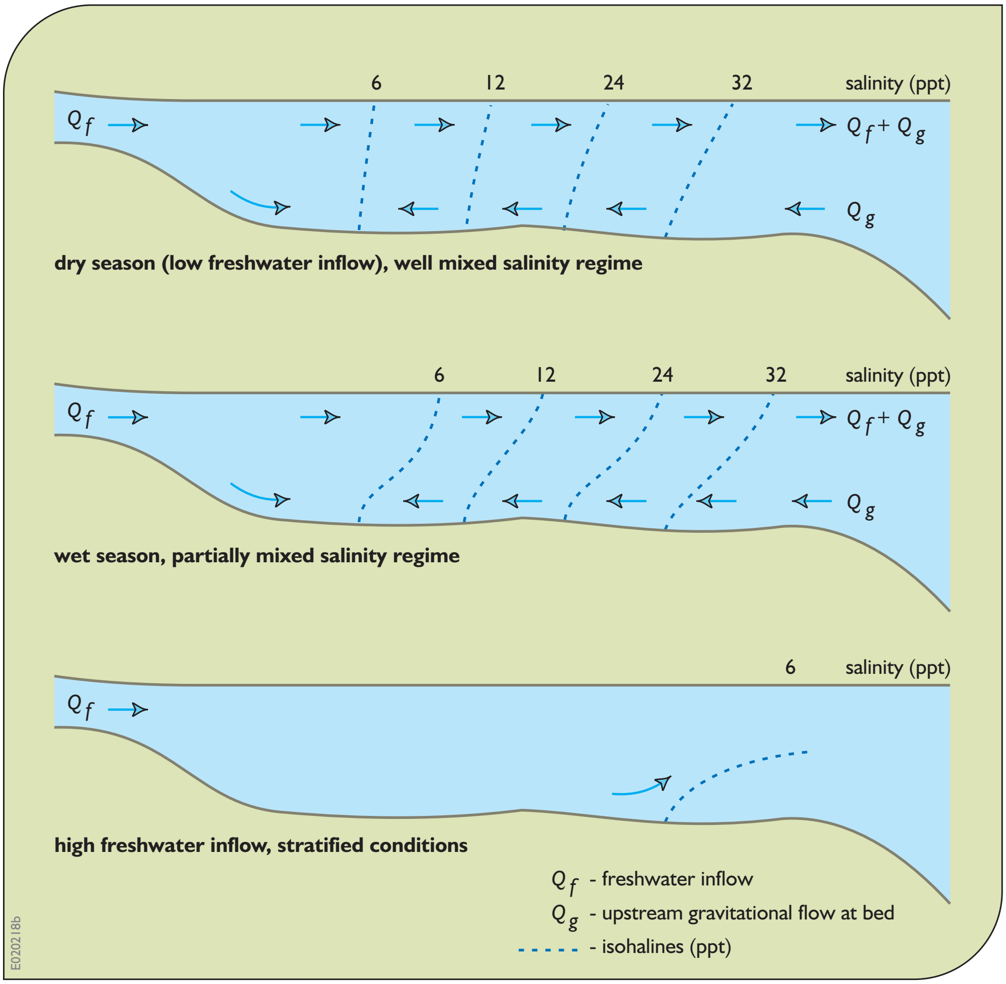



**Figure A.34.** Possible salinity regimes in tidal rivers.

the open ocean (hyper-saline conditions). During periods of wet weather, partially mixed conditions prevail. Under full flood flows, stratified conditions occur in entrance reaches and most of the salt may be washed out to sea.

## 5.7. Sediment Movement

Sediment is a major component of most estuaries. Its sources, movements and impacts are numerous.

### 5.7.1. Sources of Sediment

Different types of sediment are delivered to an estuary by a variety of sources, as illustrated in Figure A.35. Riverbank erosion and general catchment runoff produce quantities of sand, silt and clay. Catchment runoff also delivers organic matter to the river/estuary. Littoral processes in coastal waters can supply large quantities of sand to an estuary. Wind action on dunes and inter-tidal sandbanks carry fine sand into an estuary.

### 5.7.2. Factors Affecting Sediment Movement

The currents caused by freshwater inflows and tidal behaviour are the main mechanisms moving sediment in estuaries. The faster these currents, the greater will be the shear stress and turbulence generated at the bed, and the greater will be the movement of sediment by bedload and suspended load transport.

Sediment movement depends on flow, turbulence and the physical characteristics of the sediment particles. Sediment transported in water will settle at times or places of low wave or tidal activity. The rate of settling depends upon the grain size of sands and upon the mineralogy and chemistry of muds.

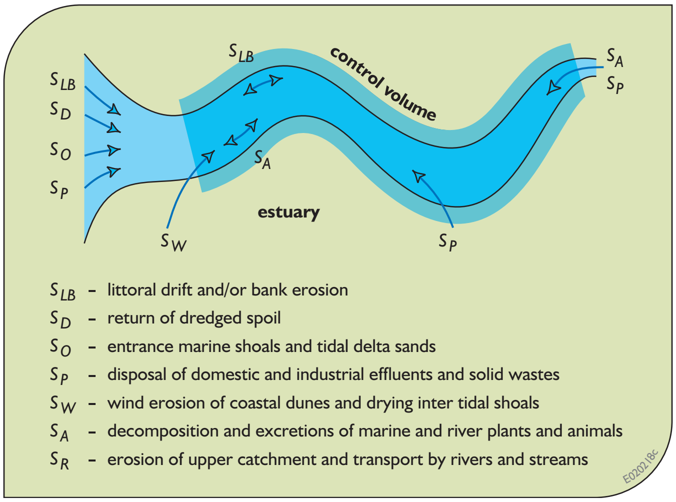



**Figure A.35. Sources of sediment in an estuary (McDowell and O'Connor, 1977).**

Sediment on the bed will be eroded and transported when the shear stress exerted on the bed by waves, tides and freshwater action, acting either alone or together, exceeds a critical minimum value. The critical shear stress also varies according to sediment size, mineralogy and chemistry.

If sediment is deposited in locations where the critical shear stress is not exceeded, or is exceeded only infrequently, then the sediment will slowly consolidate, increasing in both density and strength. As bed density increases, the stress threshold for erosion will increase, and the sediment deposit will become more stable and less likely to be eroded by natural forces.

The difference between peak flood flow and peak ebb flow velocities (tidal distortion) and gravitational circulation facilitate a net upstream movement of suspended solids, in other words, a residual suspended load flux. Because of tidal distortion, the duration of high-water slack is different from low-water slack, thereby generating differences in the slack water settling opportunity for suspended particles. This, coupled with differences in current behaviour around the two slacks, can impart a net displacement to suspended sediment each tide cycle.

Bedload transport is the principal means by which sand is moved along the estuaries. Suspended load transport of sands occurs usually in the immediate vicinity of estuary entrances, where high velocities (greater than 1 m/s) and wave action promote the suspension of sand grains.

Bedload transport is quite sensitive to small changes in velocity, such as those brought about by tidal distortion and gravitational circulation. These two mechanisms generate residual bedload flux in the upstream direction. Together, these two processes result in a net upstream transport of marine sand that forms shoals and deltas in lower estuarine reaches.

Freshwater flows during large floods can transport sand downstream at very high rates, as evidenced by the substantial scouring of shallow sand shoals that tends to occur during floods. The sand transported downstream by freshwater flows is deposited on the seaward face of the entrance bar (the ebb terminal lobe of Figure A.36). After the flood dissipates, this sand is reworked by the action of waves and nearshore currents and is returned to the shoreline over a period of many months.

The action of wind on exposed sand dunes can transport considerable quantities of sand into an estuary. Transport of sand by wind is one of the dominant factors in the sediment budget of exposed entrances and may be the principal reason for the tendency of such entrances to close.

### 5.7.3. Wind Effects

Some transport of sand grains from inter-tidal sand shoals exposed by the ebb tide also occurs through wind action. Wind-generated waves can have a significant effect on sediment production and movement along estuarine foreshores.

Within the relatively narrow confines of river estuaries, wind-generated waves are small and of short wave length (typically, the significant wave height is 0.3 m or less). Nevertheless, these waves can cause extensive riverbank erosion when the dominant wind direction coincides with a long, straight and wide stretch of the estuary.

Wind waves are often the dominant transport mechanism along the foreshores of tidal lakes (tidal and freshwater velocities in the body of the lake are generally low). Inter-tidal sand and mudflats around the peripheries of tidal lakes are formed as a result of wind waves and wind induced currents.

### 5.7.4. Ocean Waves and Entrance Effects

The movement of sand into and out of the entrance of an estuary is a complex phenomenon because of the interaction of tidal, wave and freshwater transport processes. Figure A.36 depicts the principal sediment pathways around an estuary entrance.

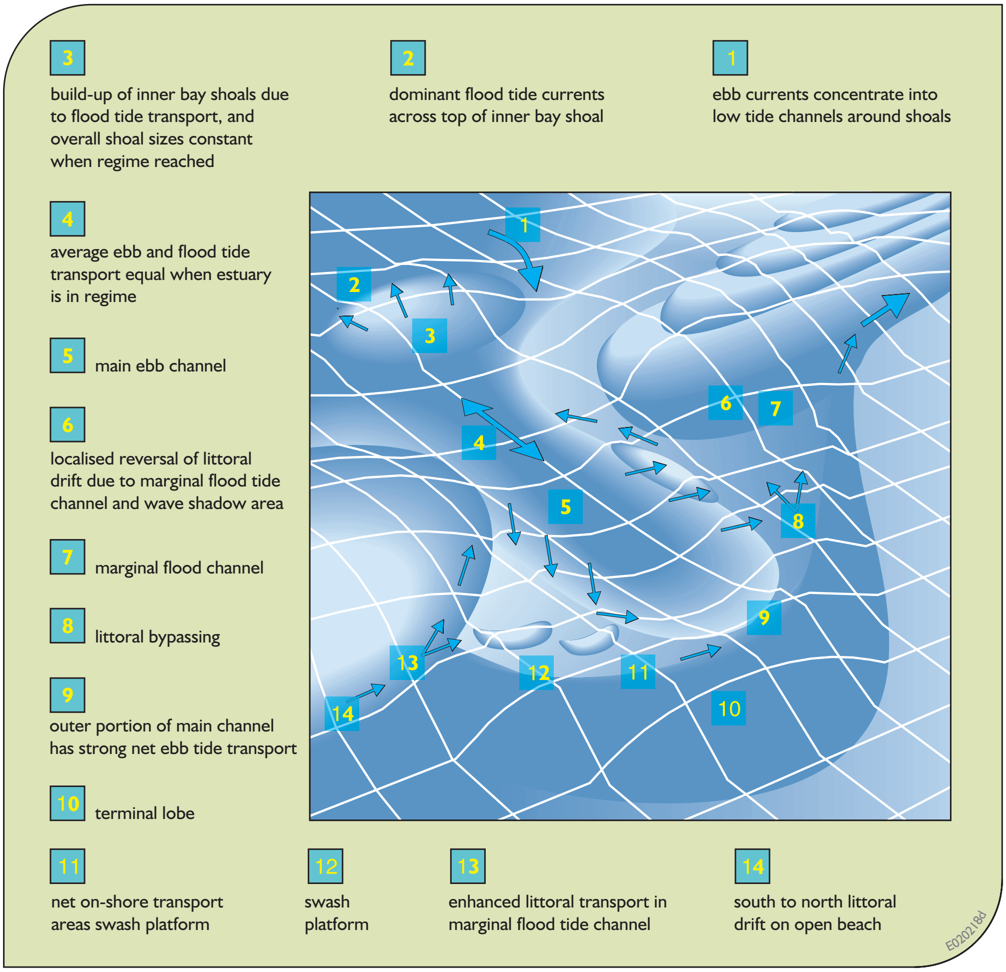



**Figure A.36.** Sediment pathways at an estuary entrance.

Waves breaking on the entrance bar bring into suspension considerable volumes of sand that can be carried into the estuary on the flood tide. Much of this sand is usually flushed back out of the estuary on the ebb tide, but a small amount may be deposited in the estuary, where it will be transported upstream by the net flood-tide transport caused by the tidal distortion.

Most of the sand transported out of the entrance by floods will be deposited on or near the entrance bar, from where it will be carried to the updrift coastline by a process known as littoral bypassing (see Figure A.36). In addition to their stirring up sediments from the entrance bar and promoting suspended sediment transport, the significance of ocean waves to estuarine sediment transport depends upon the entrance conditions.

Where there is a well-developed offshore bar, ocean waves will break on it and little wave energy will penetrate the lower reaches of the estuary. In these circumstances, ocean waves will have little effect on sediment transport within an estuary.

In contrast, ocean waves can readily penetrate the lower reaches of drowned river valley estuaries, where they combine with tidal flows to produce a relatively complex pattern of sediment movement. In some tidal lakes and rivers with relatively wide mouths, tidal sediment transport is enhanced by the penetration of waves through the entrance.

### 5.7.5. Movement of Muds

Muds are cohesive materials, and hence the processes of settling, consolidation and resistance to erosion are dependent upon inter-particle bonding and the hydrodynamic environment.

Mud particles and aggregations (flocs) of mud particles have small settling velocities. Turbulent mixing keeps mud particles in suspension throughout the tide, except at slack water when a concentrated layer of mud forms near the estuary bed. Tidal distortion and gravitational circulation result in a net upstream movement of mud particles that are deposited around the limit of the gravitational circulation current.

Muds will remain suspended in a body of moving water when the vertical mixing due to turbulence is equal to or greater than the rate of settling of individual particles. As velocities decrease, so does the vertical mixing and deposition will increase.

The settling velocity of muddy sediment is dependent upon its concentration. At high concentrations, more collisions occur between particles and the resulting flocs are larger and settle faster than the smaller particles. At low concentrations, there is little or no interaction between particles. In these cases, the settling velocity tends to be constant. At very high concentrations, the particles actively interfere with each other and hinder settling.

Like any sediments, muds begin to erode when the shear stress imposed by water movement is equal to the shear strength of the surface layer of mud. The resistance of new mud deposits to subsequent erosion increases with time. As consolidation occurs the cohesion among individual particles and flocs increases, and hence their resistance to erosion increases. Armouring also increases with time and renders the deposit more erosion resistant.

In some cases, the strength obtained through consolidation and armouring will be insufficient to resist the erosive forces of the next half-tide cycle. In these circumstances, all the material will be resuspended. In other cases, some of the deposit will be eroded and the remainder will be largely undisturbed. The residual deposit may then remain largely undisturbed until the next spring tide cycle, when it may or may not have gained sufficient strength to resist the stronger tidal currents.

The pattern of mud deposition within an estuary varies seasonally and spatially in response to the variable nature of the salinity limit. Mud particles brought into suspension are transported upstream by the gravitational circulation. This produces a turbidity maximum near the upstream limit of the circulation, leading to increased deposition of muds at slack water. In other words, the gravitational circulation leads to a trapping of muds in the estuary. The location of the zone of maximum deposition will vary with freshwater flow and the resulting pattern of salt intrusion.

The mud trapping efficiency of an estuary drops off quickly as freshwater flows increase. During major floods, the high turbulence and flushing of the flood flows is usually sufficient to keep the very high suspended load in suspension and flush it entirely out to sea. Tidal lakes, however, act as settling basins, and mud accumulates on the lake bed after each flood.

Drowned river valley estuaries and artificially deepened port areas in estuaries trap most of the mud brought into the estuary by small floods. This can be a major concern for the maintenance of shipping channels. However, during major floods most of the mud is discharged to the ocean. This is just as well, otherwise port areas might be buried in mud by a single major flood event.

### 5.7.6. Estuarine Turbidity Maximum

Most estuaries contain a specific region having an extra high concentration of suspended particles. The turbulent area in the river where the tidal forces interact with the fast-flowing fresh river waters creates a 'cloud' of suspended particulate matter (SPM): a mix of inorganic sediment, organic detritus and living organisms such as algae and zooplankton. This region is called the estuarine turbidity maximum or ETM. The ETM is very important for supporting biogeochemical, microbial and ecological processes that sustain a dominant pathway in the estuary's food web.

The turbidity maximum generally occurs near the upstream limits of the salinity intrusion due to particle trapping caused by circulation phenomena induced by tides and density-driven (gravitational) currents. Tidal shear also traps particulate matter. The turbidity maximum varies in location, spatial extent and particulate matter concentration, depending on the tidal and river flow conditions.

### 5.7.7. Biological Effects

Estuarine organisms influence both the settling and resuspension of sediments. Sea-grass beds reduce current velocities and dampen wave action near the bed, thereby increasing the settling of fine material. Filter feeders such as oysters ingest considerable amounts of fine suspended organic and mineral particulate matter that would otherwise remain in suspension. They eject waste in the form of pellets, which settle to the bottom. The activities of some other animals have the opposite effect. The feeding behaviour of many fish, birds and invertebrates disturbs bed sediments and resuspends fine particles, thereby increasing water turbidity.

The physical properties and stability of the surface layer of bottom sediments are affected by the burrowing, particle sorting and tube building activities (i.e. bioturbation activities) of benthic fauna such as crustaceans, bivalves and polychaetes. Benthic invertebrates, bacteria and diatoms produce mucus that binds sediment particles together (Madsen et al., 1993; Meadows et al., 1990; Rhoads, 1974).

These activities alter the texture, surface roughness, density, water content and shear strength of bed sediments, particularly muds (Rhoads and Boyer, 1982). Many deposit feeders ingest sediment particles at some depth and eject waste particles at the sediment surface. This waste material may consist of fine particles that are easily resuspended by currents or compacted fecal pellets that behave in a similar way to sand grains.

The remains of dead organisms can also affect sediment properties. In some areas bivalves are very abundant, and their accumulated shells form a dense layer that armours the sediment against erosive forces (Rhoads and Boyer, 1982).

## 5.8. Surface Pollutant Movement

The previous discussion on advective and dispersive transport and flushing processes that can take place in estuaries applies to dissolved substances, such as salinity and dissolved pollutants, as well as sediments. The behaviour of floating pollutants, such as oil, however, is quite different.

Oil has a very low surface tension, hence its usefulness as a lubricant. This, coupled with its hydrophobic (water-repelling) nature, means that it tends to spread over the waters surface as an unbroken thin film. Oil, when floating on a water surface, undergoes three types of transport:

- spreading
- surface-current advection
- wind-driven advection.

As oil spreads out and is transported across the surface of an estuary, the more volatile fractions evaporate and the more water-soluble fractions dissolve or emulsify with the water mass. Breaking waves facilitate emulsification. In addition to spreading out over the water surface, the oil film is carried along (advected) by surface water currents associated with freshwater and tidal discharges. Variation in surface velocities will result in additional spreading of an oil slick by lateral and longitudinal dispersion.

Finally, wind effects are of major significance to the spreading and transport of floating oil. Table A.4 shows the spread of about 500 litres of oil under a wind speed of 6–9 m/s. After three hours the spread covered 35 ha, whereas under calm conditions it was estimated that the slick would have covered only 2 ha.

The transport of some pollutants can occur in the upper surface layers of the water column (the top millimetre or so) or via surface slicks (compressed micro-layers). Organochlorines and heavy metals have been recorded in micro-layers at concentrations of up to 10,000 times greater than that which would normally occur in the water column (Hardy et al., 1990; Szekielda et al., 1972). Such concentrations may have adverse effects on planktonic organisms that gather at the sea surface. Assessing the extent, concentrations and possible impacts of micro-layers is both difficult and expensive, but their role in estuarine pollution is receiving increasing attention.

## 5.9. Estuarine Food Webs and Habitats

The tidal, sheltered waters of estuaries support unique communities of plants and animals that are adapted for life at the margin of the sea. Estuarine environments are

| time (hrs) | extent of slick (ha) |
|---|---|
| 1 | 2.5 |
| 3 | 35.0 |
| 5 | 78.0 |
| 10 | 265.0 |

E020218e

**Table A.4.** Spread of 500 litres of oil under windy conditions (wind speed 6–9 m/s) (Cormack and Nichols, 1977).

among the most productive anywhere, creating more organic matter each year than comparably sized areas of forest, grassland or agricultural land. The productivity and variety of estuarine habitats foster an abundance and diversity of wildlife. Shore birds, fish, crabs and lobsters, marine mammals, clams and other shellfish, marine worms, sea birds and reptiles are among the animals that make their homes in and around many estuaries. These animals are linked to one another and to an assortment of specialized plants and microscopic organisms through complex food webs and other interactions. For example, shore birds use their long bills to obtain fish, worms, crabs or clams. Within the mud, silt, sand or rock sediments live microscopic bacteria, a lower level of the food web, consuming decaying plants.

Many different habitat types are found in and around estuaries, including shallow open waters, freshwater and salt marshes, sandy beaches, mud and sand flats, rocky shores, oyster reefs, mangrove forests, river deltas, tidal pools, sea-grass and kelp beds, and wooded swamps.

The resident and visiting organisms in estuaries must be able to tolerate frequent and rapid changes in salinity, currents, temperatures and water levels. At high tide, seawater changes estuaries, submerging the plants and flooding creeks, marshes, mudflats or mangroves. The incoming waters seemingly bring back to life organisms that have sought shelter from their temporary exposure to the non-aquatic environment. As the tides ebb, organisms return to their protective sediments and adjust to changing temperatures.

Food webs include primary producers and primary, secondary and tertiary feeders, ranging from single-celled algae to the highly efficient predators at the top of the chain. Ecologists refer to these increments as trophic levels. The term trophic means, simply, 'pertaining to food'. The ways in which food is consumed – that is, the pattern of consumption and how it changes with time – is called trophic dynamics.

Inasmuch as an estuary is an environment characterized by rapid and frequent change, which leads to biological diversity, food webs and trophic dynamics in an estuary are complex. Unlike the open ocean, where phytoplankton species are usually the sole primary producers, estuarine systems usually contain several types of primary producers. In addition to phytoplankton, these include sea-grasses and salt-marsh plants. Zooplankton graze on phytoplankton. These, in turn, become food for plankton-eating fishes, such as herring, smelt, and the larvae and young of larger fishes. These, then, become food for carnivores and omnivores nearer the top of the food web.

Some animals graze on the larger estuarine plants, but such plants are probably more important food sources after they die and begin to decompose. Here, bacteria and fungi promote the breakdown of the dead plant material. This organic detritus is an essential source of nutrition for detritus-eating animals and supports a detrital food web.

Benthic, or bottom-dwelling, and bottom-oriented organisms are other important components in estuarine food webs. Clams, for example, reside in the bottom sediments and feed on plankton and other organic matter which they filter from the water. Oysters and mussels are other filter feeders.

Bottom deposit feeders, such as the various kinds of worms found in the estuary, move over and through bottom sediments where they find food deposited in or on the sediments. Shrimps, crabs and other invertebrates are well adapted to bottom feeding, as are many of the estuarine fishes, such as sculpin, flounder, sole and sturgeon. In fact, most fish species that reside in estuaries or move into them on feeding forays are bottom oriented in their feeding patterns.

Near the top of the estuarine food web are various carnivores and omnivores. Some show marked food preferences, while others are opportunistic feeders. These may include diving birds, wading birds, waterfowl, gulls, terns, pelicans, ospreys, trout, perch, striped bass, sharks and salmon. At the top of the food web can be seals, sea lions, whales, larger sharks and humans.

### 5.9.1. Habitat Zones

Figure A.37 illustrates various ecosystem habitats associated with an estuary. Shown on the figure are nine different habitats.

Fish habitats are mainly underwater. Although not readily visible, they play an important role in an estuary and in the economy of the region. There are many different types of habitats on which fish depend for their survival, ranging from deep ocean water to the shallower areas near the coast to rivers and lakes far inland. Like one

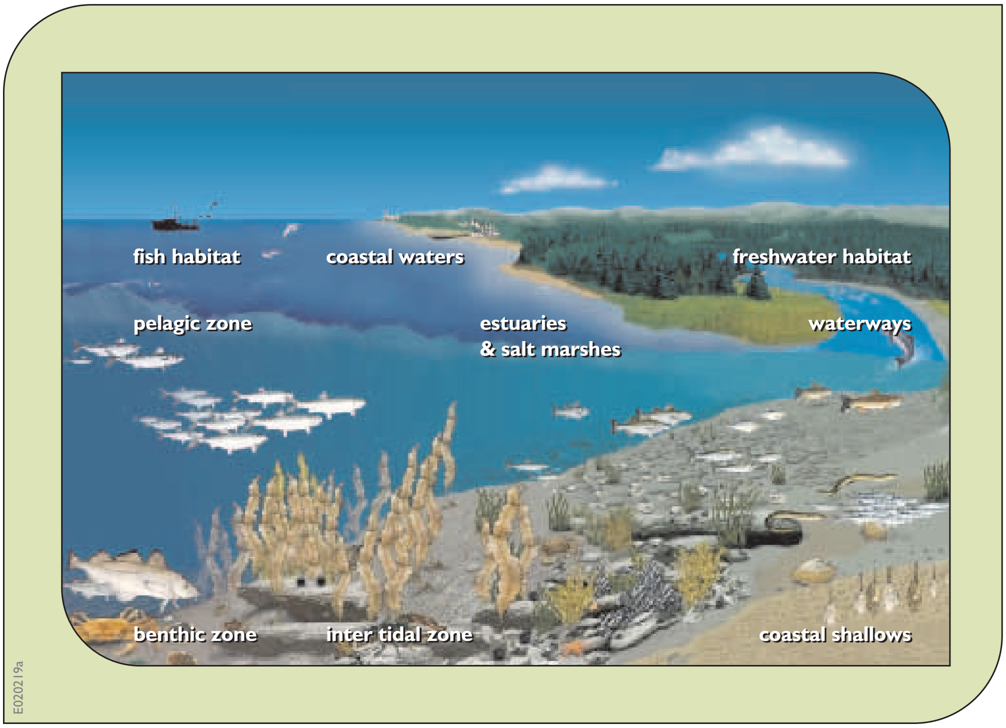



**Figure A.37.** Typical habitats of an estuary and its surroundings.

giant mosaic, habitat pieces are linked together and support one another. Habitat connectivity is critical. What affects one habitat will in turn eventually affect other, connected habitats.

Coastal waters can support commercial and recreational fishing industries. They can also support a variety of aquaculture activities such as shellfish culture or fish farms. In many communities, fishing is an important contributor to the regular diet of the local population.

Freshwater habitats are not much different from their marine counterparts in supporting a wide and diverse fish population. Lakes provide deep and stable habitats of rocky, sandy and muddy bottoms. Aquatic plants provide food, shelter and spawning areas, as they do in the oceans and estuaries. Unpolluted rivers and streams can offer a habitat of clean, well-oxygenated water. Insects living and falling in the water are a major food source for fish. Fish use the rivers to travel between spawning, rearing and feeding habitats.

Pelagic fish such as mackerel, tuna and herring that live their entire lives just below the surface of the ocean often follow the ocean currents, which are like rivers in the ocean. The fish habitat in these areas can change rapidly due to the actions of the wind, rain, sun and air temperature. The fish follow these 'rivers in the ocean' to areas that are suitable for them to reproduce and feed. They lay their eggs in the open sea or shallow coastal waters. When the eggs hatch, the young drift with the ocean currents. Typically, out of millions that hatch from eggs, only a few survive to become adults.

Many nutrients brought down from streams and rivers pass through estuaries and salt marshes before reaching the oceans. These are among the most fertile areas found anywhere for plants and animals. Coasts that contain extensive estuaries and salt marshes support the fisheries on which many depend. The nutrients fertilize large expanses of the plants along shores. These are the main feeding, growing and sheltering areas for young fish and shellfish. After hatching in freshwater, young smelt drift downstream to the estuaries where they hide and feed among aquatic plants.

Cod, halibut and other fish that live on or near the bottom of the ocean are called 'benthic fish' or 'groundfish'. Others that are harvested from these deep-water areas can include snow crabs and scallops. In the waters nearer to the shore, the fish habitat is a mixture of rock and mud. This near shore mixture can be the home to lobsters, rock crabs, shellfish and smaller fish. The physical and chemical conditions in these areas typically do not change very quickly. If any sudden changes do happen, the effects on the fish habitat and the fish themselves can be severe.

Shallow water habitats are usually a mixture of mud, rocks and gravel. Animals like oysters, clams and mussels that filter the water for food can find abundant food in this habitat. They feed on microscopic animals and plants called 'plankton'. They spend their lives in one place, so their habitat is extremely important for shelter and feeding. Urban or commercial developments along coasts, poorly designed wastewater treatment systems, sewage outfalls and infilling of marshes and bays can all degrade the ecosystem habitat in these areas. The filter-feeding animals in these areas are very sensitive to changes in their habitats such as a lack of oxygen, increases in temperatures or being buried by mud.

In the shallower waters and rocky shores, the increased temperature, sunlight and wave action results in more types of habitats that support more kinds of plants and animals. Plants become very important in this area, transferring the sun's energy through plant-eating animals. This inshore area supports a wide variety of animals that use the plants for food, refuge, hunting grounds and reproduction. For example, periwinkles, urchins and starfish scrape the algae off the rocks and plants, and also lay their eggs on them. These eggs are sources of food for crabs, lobsters and other fish.

## 5.10. Estuarine Services

Besides serving as important habitats for wildlife, estuaries and the wetlands that often fringe them perform other valuable services. Water draining from the uplands carries sediments, nutrients and other pollutants. As the water flows through fresh and salt marshes, much of the sediment and pollution is filtered out, creating cleaner and clearer water that benefits both people and marine life. Wetland plants and soils also form a natural buffer zone between the land and ocean, absorbing floodwaters and dissipating storm surges. This helps protect upland organisms as well as valuable real estate from storm and flood damage. Salt marsh grasses and other estuarine plants also help prevent erosion and stabilize the shoreline.

Among the cultural benefits that estuaries offer are recreation, a source of scientific knowledge, education and aesthetic aspects. Boating, fishing, swimming, surfing and bird watching are just a few of the recreational activities people enjoy there. Estuaries are often the cultural centres of coastal communities, serving as the focal points for local commerce, recreation and culture. They also provide aesthetic enjoyment for the people who live, work or holiday in and around them.

Finally, estuaries provide tangible and direct economic benefits. Tourism, fisheries and other commercial activities thrive on the wealth of natural resources estuaries supply. The protected coastal waters of estuaries also support important public infrastructure, serving as harbours and ports vital for shipping, transportation, and industry.

To maintain and enhance these and other services and benefits derived from estuaries, they must be managed. This often includes protection and restoration.

## 5.11. Estuary Protection

The economy of many coastal areas is based primarily on the natural beauty and bounty of estuaries. When those natural resources are imperilled, so too are the economies and well-being of people who live and work there. In many countries that have coastlines, most of their populations are concentrated along them. Populations in coastal regions are growing faster than in inland regions throughout most of the world.

Population pressures and the impact of human encroachment on estuaries, as shown in Figure A.38, is evident. In North America, for example, some 70% of the original sea-grass meadows and salt marshes have been lost in Puget Sound, Galveston Bay and Narragansett Bay. Over 90% of original wetlands have been lost in San Francisco Bay, Chesapeake Bay, Hudson-Raritan Estuary, and Tampa Bay (USEPA, 2001).

This increasing concentration of people living along coastlines tends to change the natural balance of estuarine ecosystems and threatens their integrity. Channels are dredged, marshes and tidal flats are filled, waters are polluted, and shorelines are reconstructed to accommodate human housing, transport and recreational needs. Stresses caused by overuse of resources and unchecked land use practices in some areas have resulted in unsafe drinking water, closing of beaches and shellfish beds, harmful algal blooms, unproductive fisheries, loss of habitat, death of fish and wildlife, and a host of other human health and natural resource problems.

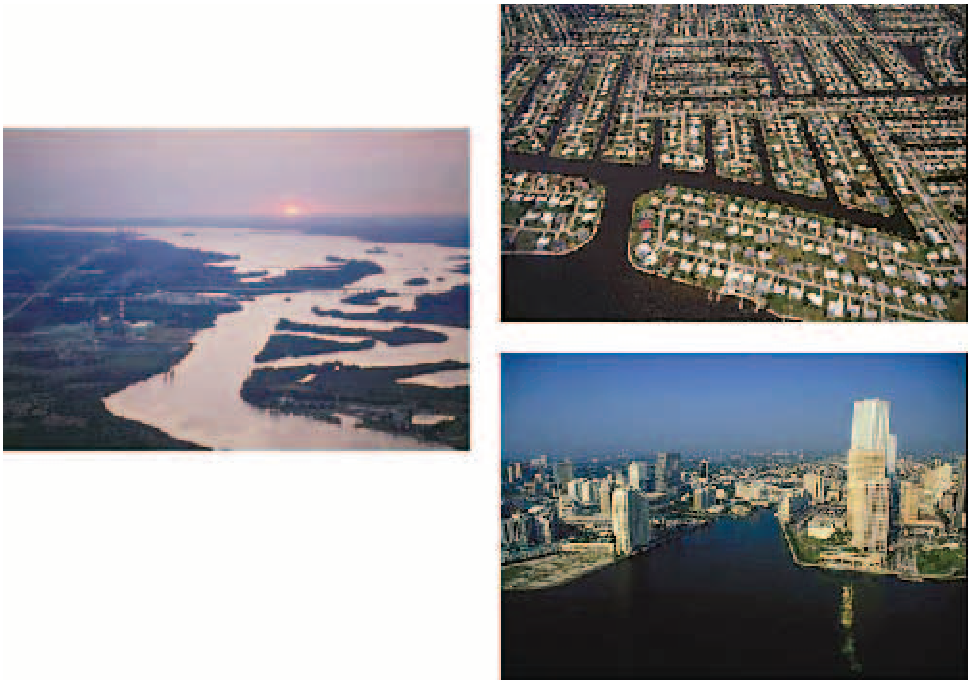


**Figure A.38.** Not all estuaries remain natural. Some show the impacts of development and the desire for humans to live near and on them. (South Florida Water Management District, West Palm Beach, Florida.)

Polluted runoff from rural, suburban and urban areas upstream of estuaries can also cause damage. Stormwater can pick up contaminants from roads, vehicles, lawns and construction sites, and discharge them into streams, lakes and wetlands. Everything that happens on land in the watershed that drains into the estuary can affect its habitat. When stream channel habitat is degraded, say from agriculture, construction or forestry activities, fish die because their nesting and feeding areas are destroyed. In urban harbours, in particular, polluted runoff creates 'hot spots' of toxic contamination where relatively few species can live. For example, polluted runoff from agricultural chemicals is responsible for the 'dead zone' off the coasts of Texas and Louisiana, which extends some 11,000 km$^2$ into the Gulf of Mexico. Airborne pollution that falls on estuaries also contributes to their contamination.

Dams blocking river flows can restrict the upstream and downstream passage of migrating fish, isolating them from spawning and feeding areas. The construction of dams accounts for significant and ongoing loss of habitat in the watersheds of many of the world's rivers and estuaries. Without healthy streams, estuaries cannot receive their normal allocation of nutrients. The result is usually reduced productivity. If fewer fish return to the estuary, many living organisms in the food web that depend on healthy fish populations will suffer.

Excessive discharges of pollutants in wastewater effluent from city and industrial sewage treatment plants can degrade estuary ecosystems that require clean water to survive and thrive. This is evident in many if not most of Western Europe's major estuaries; examples include the Tyne, Tees, Humber, and Thames Estuaries in the UK; the Scheldt Estuary in Belgium and the Netherlands; the Ems Estuary in Germany and the Netherlands; the Weser and Elbe Estuaries in Germany; and the Seine Estuary in France. Wastewater treatment can help restore some of the damage done by excessive pollution. When the treatment plants discharging into Tampa Bay (Florida) were upgraded to advanced technology, the sea-grasses – 85% of which had been destroyed – began to grow back, and along with them the fish and other creatures that depend on those grasses. Estuaries can be restored, but it takes time as well as money.

Shifts in climate and the altering of stream channels by humans cause an annual loss of tens of thousands of hectares of estuary habitat. This is of concern along the Louisiana coast of the United States. Changing the course of the Mississippi River as it enters the Gulf of Mexico along the coast of Louisiana has substantially reduced the yearly input of sediment to its delta. The result has been an open-sea encroachment of both the estuaries and dry land. Louisiana estuaries lose about 12,000 hectares of land each year to erosion and subsidence (actual sinking of the land into the water). Delta erosion and loss has also been a problem in the River Nile Estuary since the building of the High Aswan Dam reduced the sediment loads normally carried by the Nile River to the Mediterranean Sea.

Estuarine habitat and freshwater loss is a major challenge to estuarine management throughout the developed and developing world. Population growth and economic development will almost certainly lead to additional loss of habitat and freshwater diversion in many estuaries. The same social driving factors, combined with a continuing migration of human populations to coastal urban centres, require major expansions of engineering infrastructure. Nutrient emissions from human activities can greatly exceed even those naturally carried by large rivers. An increasing number of coastal areas are already manifesting serious effects of nutrient over-enrichment, including bottom-water hypoxia or anoxia, undesirable algal blooms, and the loss of sea-grasses and coral reefs.

## 5.12. Estuarine Restoration

The most cost-effective route to saving estuaries is to prevent habitat alteration and destruction in the first place. But because of the substantial loss of vital estuarine habitat, and the habitat that continues to be destroyed, restoration is often necessary.

'Restoration' means returning an area of estuary habitat to a successful, self-sustaining ecosystem with both clean water and healthy habitats that support fish and wildlife and human uses of the estuary, such as swimming, boating and recreational and commercial fishing and port activities. Ecological restoration does not usually focus on a single species but strives to replicate the original natural system. The goal is to help rebuild a healthy, functioning system that works as it did before it was polluted or destroyed. Restoration also means an actual increase in the quantity of high-quality estuary habitats, as measured both by their surface area and by their ability to support a resilient healthy ecosystem.

Restoration activities in estuaries range from the simple to the complex. They may include, singly or in combination:

- baseline assessments
- setting performance standards and defining long-term monitoring and conservation plans
- restoring the physical and hydrological conditions through engineered activities, often involving heavy equipment and the returning of tidal waters
- reducing inflows of nutrients, BOD and other pollutants
- chemical cleanup of toxic substances
- revegetation of an area through native plantings or natural regrowth.

Restoration of an estuary is most effectively done by communities that live in the watersheds that impact that estuary. These communities should take a watershed approach to estuary restoration. Estuaries are often at the heart of local economies and traditions. Although they may need additional financial and technical assistance from federal, state and local governments, the people who live near estuaries are the ones who will determine just how successful any restoration effort will be.

## 5.13. Estuarine Management

To achieve estuarine protection and restoration and all the benefits that can be derived from them, estuaries must be managed. Management must begin with some attention to the impacts of daily activities of people who live in watersheds that drain into the estuaries. These activities can affect the ecological habitat of waterways and estuaries. Everything that enters into a stream or river or any other type of waterway that drains into an ocean will eventually find its way to the estuaries and oceans, and thus can eventually affect the estuarine environment. Pollution or even mud runoff in the water can affect aquatic habitats downstream and interfere with plant and animal reproduction. For example, contaminants washing downstream can cause the closure of shellfish beds for human consumption or kill large numbers of fish.

Although each estuary is unique, many face similar environmental problems and challenges that include urban and commercial development, over-enrichment of nutrients, pathogen contamination, toxic chemicals, alteration of freshwater inflow, loss of habitat, declines in fish and other wildlife, and introduction of invasive species.

The following discussion provides a brief overview of the most common problems facing estuary management and restoration. These problems are primarily caused by human activities.

### 5.13.1. Engineering Infrastructure

Many engineering developments such as harbours, training walls, navigation channels, reclamation and dredging typically take place in the mouths of estuaries. By virtue of their ability to alter depths significantly in the most tidally sensitive reach, such developments can affect tidal behaviour along the entire estuary. Often complex hydrodynamic-ecological models are needed to estimate the impacts of such infrastructure.

### 5.13.2. Nutrient Overloading

Nutrients such as nitrogen and phosphorus are necessary for the growth of plants and animals, and to support a healthy aquatic ecosystem. In excess, however, these nutrients can contribute to fish disease, red or brown tide, algae blooms and low levels of dissolved oxygen. The condition where dissolved oxygen is less than 2 mg/l is referred to as hypoxia. Many species are likely to die below that level. The concentration of dissolved oxygen in healthy waters is at least 5 or 6 mg/l.

Nutrients can come from both point sources and non-point sources such as sewage treatment plant discharges, stormwater runoff from lawns and agricultural lands, faulty or leaking septic systems, sediment in runoff, animal wastes, atmospheric deposition originating from power plants or vehicles, and groundwater discharges. Excessive nutrients stimulate the growth of algae. As the algae die they decay, and this decreases the dissolved oxygen in water. Algae also reduce sunlight penetration in the water. Fish and shellfish deprived of oxygen, and underwater sea-grasses deprived of light, will die. Animals that depend on sea-grasses for food or shelter leave the area or die. In addition, the excessive algal growth may result in brown and red tides that have been linked to fish kills, manatee deaths and negative impacts on scallops. Decaying algae may also cause foul smells and decreased aesthetic value.

### 5.13.3. Pathogens

Pathogens such as viruses, bacteria and parasites, as well as certain types of algae, can be toxic and may cause diseases. When found in marine waters they can pose a health threat to swimmers, surfers, divers and seafood consumers. Fish and filter-feeding organisms such as shellfish concentrate pathogens in their tissues, and people eating them may become ill. Pathogen contamination can result in the closure of shellfishing areas and bathing beaches.

Sources of pathogens include urban and agricultural runoff, boat and marina waste, faulty or leaky septic systems, sewage treatment plant discharges, combined sewer overflows, recreational vehicles or campers, illegal sewer connections, and waste from pets or wildlife.

### 5.13.4. Toxic Chemicals

Toxic substances such as metals, polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), heavy metals, and pesticides are a concern in the estuarine environment. These substances can enter waterways through stormdrains; discharges from industry and sewage treatment plants; runoff from lawns, streets and farmlands; and deposition from the atmosphere. Many toxic contaminants are also found in sediments and are resuspended into water by dredging and boating activities. Bottom-dwelling organisms are exposed to these chemicals and if consumed may pose a risk to human health. As a result there may be fishery and shellfish bed closures, and bans on the consumption of certain foods.

### 5.13.5. Habitat Loss and Degradation

The continued health and biodiversity of marine and estuarine systems depends on the maintenance of high-quality habitats. The same areas that often attract human development also provide essential food, cover, migratory corridors and breeding/nursery areas for a broad array of coastal and marine organisms. In addition, these habitats perform important functions such as enhancement of water quality and flood protection, and water storage.

Estuarine ecosystems can be degraded through loss of habitat – such as the conversion of a sea-grass bed to an island of dredged material – or through a change or degradation in structure, function or composition. Threats to habitat include commercial and residential development, highway construction, marinas, dyking, dredging, damming and filling. Wetland loss and degradation caused by dredging and filling reduces the amount of habitat available to support healthy populations of wildlife and marine organisms. All of these activities may result in an increase in the runoff of sediments, nutrients and chemicals.

### 5.13.6. Introduced Species

Intentional or accidental introduction of invasive species can often result in unexpected ecological, economic and social damages. Through predation and competition, introduced species have contributed to the eradication of some native populations and substantially reduced others, fundamentally altering the food web. Overpopulation of some introduced herbivores has resulted in overgrazing of estuarine vegetation and the resultant degradation and loss of marsh. Other impacts can include:

- alteration of water tables
- modification of nutrient cycles or soil fertility
- increased erosion
- interference with navigation, agricultural irrigation, sport and commercial fishing, recreational boating, beach use
- possible introduction of pathogens.

Sources of introduced species include ship ballast, mariculture and aquarium trade.

### 5.13.7. Alteration of Natural Flow Regimes

Alteration of the natural flow regimes of tributaries can have significant effects upon water quality and the distribution of living organisms within estuaries. Freshwater is an increasingly limited resource in many areas. Human management of this resource has altered the timing and volume of inflow to some estuaries. Too much or too little freshwater can adversely affect fish spawning, shellfish survival, bird nesting, seed propagation, and other seasonal activities of fish and wildlife. In addition to changing salinity levels, inflow provides nutrients and sediments that are important for overall productivity of the estuary.

### 5.13.8. Declines in Fish and Wildlife Populations

The distribution and abundance of estuarine fish and wildlife depend on factors such as light, turbidity, nutrient availability, temperature, salinity and food availability. Natural and human-induced events that disturb or change environmental conditions affect the distribution and abundance of estuarine species. Declines in fish and wildlife populations have resulted from fragmentation and loss of habitats and ecosystems, pollution and decreased water quality, over-exploitation of resources, and the introduction of exotic or nuisance species.

Habitat loss and degradation can lead to decreases in the stocks of sport and commercial fish and shellfish, changes in the populations of fur-bearing and waterfowl species, and decreasing habitat for migratory birds and other species. Pollutants such as herbicides, pesticides and other wastes pose a threat to living resources by contaminating the food chain and eliminating food sources. Runoff from farms and cities, and toxic releases, can alter aquatic habitat, harm animal health, reduce reproductive potential, and render many fish unsuitable for human consumption.

Other threats to wildlife include oil spills, bioaccumulation of toxins, outbreaks of contagious and infectious diseases, and algal blooms such as red and brown tides. Over-exploitation occurs when fisherman, trappers, hunters or collectors take so many individuals of a species that its ability to maintain stable population levels is impaired. Introduced species compete with native species for food and habitat. Other causes of decline in fish and wildlife populations include agricultural and logging activities, trawling, boat disturbances, entanglement from marine debris, and changes in freshwater inflow.

## 6. Coasts

Many factors threaten the resilience of naturally dynamic and generally fragile coastal systems. Most originate from land development and urbanization, pollution, overuse and accelerated sea level rise. The challenge in coastal zone management is to maintain the natural physical variability and ecosystem diversity present in, and all the resulting benefits derived from, coastal systems, while simultaneously satisfying multiple stakeholder interests and competing resource uses and values that often favour the stabilization of such zones.

### 6.1. Coastal Zone Features and Processes

The coast is where land and ocean interact. It is a zone extending inland to the limit of tidal or sea spray influence. The relative levels of the sea and land determine coast locations. In relation to the land, sea levels fall and rise as glaciers grow and melt over periods of thousands of years. When sea levels rise, as is being observed today, the coasts move inland unless prevented from doing so by barriers. Most efforts to 'protect' the coasts are oriented toward preventing this inland movement and destruction of property – property that is often very desirable and hence very valuable. But coastal beaches are naturally dynamic, adjusting themselves, through erosion and deposition, to the variable and often random forces they are subjected to. Coasts are shaped by land and marine processes, which in turn are increasingly being influenced by human activities and structures that tend to constrain these natural dynamic processes. Preventing erosion, for example, will eventually destroy the beaches, and beaches provide much more than just recreational benefits, as will be discussed below.

Land processes have primarily shaped the irregular, glacially scoured, rocky coasts typically found in higher latitudes. These are characterized by long narrow embayments carved into bedrock. Other coasts have a combined marine and land origin, or a primarily marine one. The Outer Banks of North Carolina on the US Atlantic coastline are barrier islands formed by the rising sea. The Pamlico and Albemarle Sounds behind them, on the other hand, are simply a series of flooded river valleys. The Mississippi Delta comprises sediment that is transported by the Mississippi River into the Gulf of Mexico, where it is sculpted by waves into long fingers of land protruding into the sea. The same is true of the Nile River Delta in Egypt.

Coasts formed by marine processes include much of the US Pacific Coast, with its cliffs formed by wave erosion. Outer Cape Cod, jutting into the Atlantic Ocean in Massachusetts, is another coast formed by the wave erosion of a deposit of glacial sands and gravel. A predominantly marine-formed coast in North America extends from Long Island, New York, to Mexico's Yucatan, with only a few interruptions. These barrier islands just offshore of the eastern coast of the United States occupy more shoreline distance than any other open-ocean shoreline type. The islands are formed by the waves, the wind and the offshore currents.

### 6.1.1. Water Waves

Every coastal project such as beach nourishment or harbour design requires information on the wave conditions in the region of interest. Increasing demands for accurate design wave conditions and for input data for the investigation of sediment transport and surf-zone circulation have resulted in improved wave-prediction models during the last two decades. Liu and Losada (2002), Dean and Dalrymple (1984), Dingemans (1997), Mei and Liu (1993), Horikawa (1978), Ippen (1966), Kinsman (1965), and Sarpkaya and Isaacson (1981) discuss various theoretical wave models and cite original papers.

In spite of model improvements, the prediction of waves and their effects is still relatively primitive in comparison with the complexity of the real system. Numerical models are still based on simplified governing equations, boundary conditions and numerical schemes, imposing different restrictions to practical applications. The computational effort required to solve a truly three-dimensional wave propagation problem, involving numerous physical processes resulting in hundreds or more wavelengths with different temporal and spatial scales, is still too large to be feasible in engineering design practice at this time.

Local currents, accelerations and pressure fluctuations often accompany water waves. Their simplest form is sinusoidal (Figure A.39).

Small amplitude wave theory is based on the sinusoidal wave shown in Figure. A.39. A right-hand coordinate system is commonly used, with its origin at still-water level. Still-water level, *SWL*, is the water surface that would exist in the absence of any wave action. The *x*-axis is horizontal and parallel to the direction of wave propagation and no variation is assumed in the y direction, perpendicular to the *x*-axis.

The sinusoidal water surface level, $\eta$, at any time $t$ at location $x$ may be described by:

$$\eta = a\cos\left(\frac{2\pi x}{L} - \frac{2\pi t}{T}\right) \tag{A.2}$$

where $a$ is the amplitude of the wave, $x$ is distance in the direction of wave propagation, $t$ is time, $L$ is the wavelength, $\pi$(pi) radians is 180° and $T$ is the wave period.

The maximum vertical distance between crest and trough of the wave is the wave height, $H$ ($=2a$). The distance over which the wave pattern repeats itself is the wavelength, $L$. The waves propagate with a velocity, $C$, and the time required for a wave to pass a particular location is the wave period $T$ ($T = L/C$). The inverse of the wave period is the wave frequency $f$ ($f = 1/T$). The

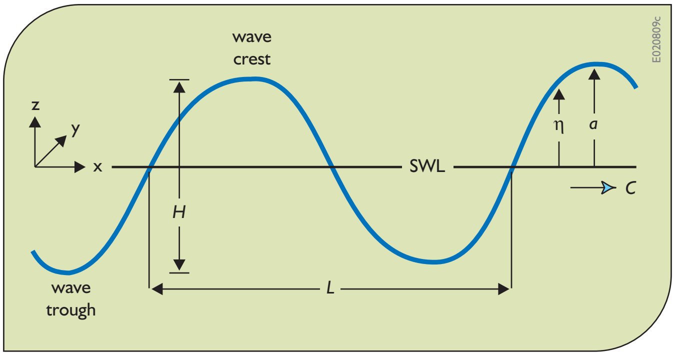



**Figure A.39. Basic wave components and characteristics include wave height, *H*, length, *L*, amplitude, *a*, relative to still-water level, *SWL*, and velocity *C*.**

ratio of wave height to wavelength ($H/L$) is called wave steepness.

The high water levels are the wave crests; the low levels are the wave troughs. The crest of waves is measured in the $y$ direction. Swell is normally characterized by long crested waves, each extending in the $y$ direction over a hundred metres or so. Sea normally has short-crested waves with local peaks.

Finally, waves are said to be in deep water when the depth of the water per unit wavelength, $L$, is greater than 0.5 and in shallow water when it is less than 0.05. Between these conditions, the water depth is called transitional.

Water waves range from capillary waves that have very short wave periods (order 0.1 second) to tides, tsunamis (earthquake generated waves) and seiches (basin oscillations), where the wave periods are expressed in minutes or hours. Waves also vary in height, from a few millimetres for capillary waves to tens of metres for the long waves. Gravity or wind-generated waves are in the middle of that range. They have periods from 1 to 30 seconds and wave heights that are seldom greater than 10 m and mostly of the order of 1 m. They are generated by wind against the gravitational force attempting to restore the still-water level. These gravity or wind-generated waves account for most of the total available wave energy.

The actual shape of a water surface subjected to wind does not look like Figure A. 39. It is a complex combination of many waves of many different wave heights and periods. These waves propagate more or less in the wind direction. Local peaks in the water levels occur when and where the two waves combine, and lower water levels occur when and where the waves separate, resulting in the irregular wave pattern at any particular time and location. Even when the first puffs of wind touch a flat water surface, the resulting distortions present non-linearities that make rigorous analysis extremely difficult, if not impossible. When the first ripples generated by these puffs are subsequently strengthened by the wind and interact with each other, the result is known as a confused sea. As the wind speed and duration increase, the complexity of the waves will increase. Coastal zone modellers need to understand this confusion.

Wind waves generated by a storm over the sea or ocean usually consist of a wide range of wave frequencies. Wave components with higher wave frequencies propagate at slower speeds than do those with lower ones. As they propagate towards the coast, long waves lead the wave group and are followed by short waves. In deep water, wind-generated waves are not affected by the bathymetry. Upon entering shallow waters, however, they are either refracted by bathymetry or current, or diffracted around abrupt bathymetric features such as submerged ridges or canyons. A part of wave energy is reflected back to the deep sea. Continuing their shoreward propagation, waves lose some of their energy through dissipation near the bottom. Nevertheless, each wave profile becomes steeper with increasing wave amplitude and decreasing wavelength.

In very shallow water the front face of a wave moves at a slower speed than the wave crest, causing the overturning motion (the 'breaking') of the crest. Such an overturning motion usually creates a jet of water, which falls near the base of the wave and generates a large splash. Turbulence associated with breaking waves is responsible not only for the energy dissipation but also for the sediment movement in the surf zone.

On large bodies of water, waves can travel well beyond the area in which they are generated. For example, waves generated by a storm off the coast of Newfoundland in Canada may travel easterly toward Europe, eventually arriving with much less energy on the coast of Portugal. Wave energy frequencies are reduced. The resulting waves arriving in Portugal, for example, will be more orderly than the initial sea generated off Newfoundland, with longer wave periods (10–20 seconds), smaller wave heights and more pronounced wave grouping. Waves generated some distance away are called swell.

### 6.1.2. Tides and Water Levels

Water levels influence both flooding and wave exposure. Most damage to structures along coastlines occurs when the water levels are high. Rising water levels expose shore structures to larger waves because the water depth determines where waves break and thus where they release most of their energy. The increased forces and overtopping of water may damage coastline structures and areas behind it. Conversely, when the water level drops, the same structures may not be exposed to waves at all.

Similarly, high water levels can cause sandy shores to retreat, even if the shores are backed by substantial dunes. Higher water levels allow larger waves to come closer in to the shore. These waves erode dunes and upper beach and deposit the sand offshore. If the water level rise is temporary, most of this loss will be regained at the next low water. Permanent water level rise, however, will result in permanent loss of sand. Shorelines consisting of bluffs or cliffs of erodible material, such as glacial till or soft rock, are continually eroded by wave action. High water levels, however, will allow larger waves to attack the bluffs directly, increasing the rate of shoreline recession.

### 6.1.3. Coastal Sediment Transport

Although there are many important aspects of coastal zone management, the most important driver and fundamental criterion in a coastal zone management plan is often the movement of sand and other sediment.

The transport of sediment, moved by waves and wind, may be divided into cross-shore and long-shore components. Sediment movement can result in erosion or accretion: removal or addition of volumes of sand. Erosion normally results in shoreline recession (movement of the shoreline inland). Deposition of sediments on beaches (accretion) causes the shoreline to move out to sea. Most protection schemes do not function well with too much cross-shore sediment movement. In particular, if the main cause of shoreline recession is systematic movement of sand offshore, protection becomes difficult.

### 6.1.4. Barrier Islands

Barrier islands, as illustrated in Figure A.40, are ribbons of sand formed by rising sea levels on coastal plains with low, flat surfaces extending to the edge of the sea, a large supply of sand, and waves large enough to move the sand about. Barrier island migration towards the mainland occurs as the barrier island rolls over itself like a wheel. The ocean side of the island retreats as storms push sand across the island to form sand overwash fans. Overwash fans often extend into the lagoon behind the island and may cause the island to widen in a landward direction. Simultaneous shoreline retreat of the side of the island open to the ocean results in island migration.

### 6.1.5. Tidal Deltas and Inlets

Inlets usually form during storms when stormwaters that have been forced into estuaries behind a barrier island rush across the island on their return to the sea, cutting a channel through the island. Figure A.40 illustrates a natural inlet.

Tidal deltas, as illustrated in Figure A.40, are bodies of sand formed at inlets by tidal currents. Flood tidal deltas are formed by sand forced into the lagoon between barrier islands and the mainland by incoming or flood tides. The ebb tidal delta is the body of sand pushed seaward by the outgoing or ebbing tide. Salt marshes or mangroves, where mangroves grow and eventually add to the width of the barrier islands, colonize flood tidal deltas. Overwash fans also add to the width of barrier islands once the inlet migrates or opens to the sea.

When an inlet is jettied to stabilize a navigation channel, the ebb tidal delta is broken up. For a few years or decades after jetty construction, the sand from the ebb tidal delta ends up being deposited on adjacent shorelines. Eventually a new and completely submerged tidal delta forms far offshore at the end of the jetties, taking sand permanently away from the beach.

### 6.1.6. Beaches

Beaches are strips of unconsolidated material found at the seaward margin of the coast. The location of the beach shoreline, the wet–dry boundary of the beach, moves up and down with the tide. (Maps usually show the mid-tide line or mean sea level.) A beach on a barrier island or spit is called a barrier beach, and beaches between rocky headlands are called pocket beaches.

Beach materials are subjected to wind, waves and currents, which shape the beaches accordingly. They

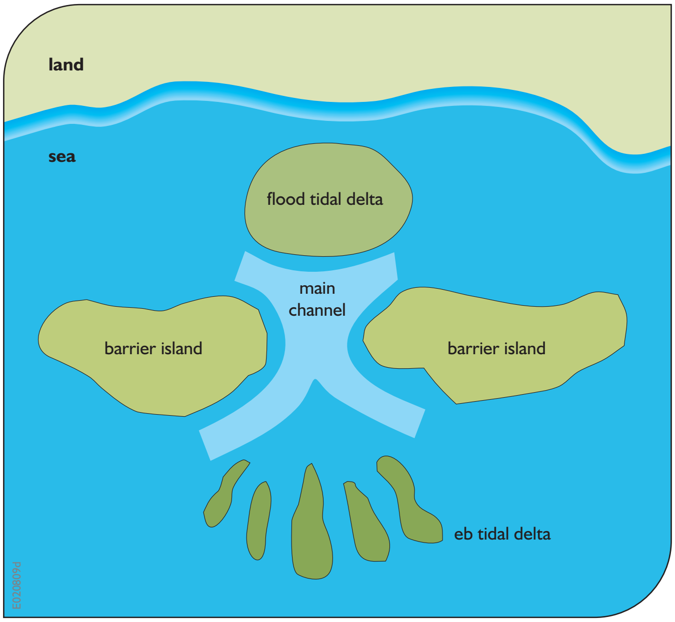



**Figure A.40.** Barrier island main channel inlets cut through a barrier island and tidal deltas.

constantly adjust to wave and tidal energy, the quality and quantity of the sediment supply, and level of the sea. One of the challenges confronting coastal zone managers is to predict the future behaviour or movements of beaches.

Beaches can be classified by the size of the material on them (e.g. mud, sand, gravel, boulders), or by the type of material (e.g. volcanic, coral, quartz). Beaches can also be referred to as high energy, moderate energy or low energy beaches according to some average regional wave height.

Storms dramatically increase wave energy, which is proportional to wave height, as well as a rise in sea level. Beaches adjust their shapes in response to this higher energy environment. Sand will be moved, offshore bars may form, or the beach may flatten out. The storm-related rise in sea level helps waves reach dunes or bluffs, washing dune sand back to the beach. This addition of sand and sediment on the beach further buffers the waves and wind. The amount of sand movement and the ultimate shape taken by the beach will depend in part on this sediment supply and the grain size of the beach material.

During a storm the beach flattens as sand moves from the upper to the lower beach. This dissipates wave energy over a broadened surface relative to the beach before the storm. After the storm the sand will gradually return to nearly its pre-storm profile.

As shown in Figure A.41, beaches are steeper during calm weather, and flatter during storm events. All beaches with a reasonably good sand supply and no seawall will respond to strong storm waves by either flattening or forming offshore bars. Beaches that flatten in response to large waves cause breaking waves to dissipate their energy over a broader surface. Waves that reach cliffs before having their energy dissipated must dissipate their energy over a relatively short distance. The result of course is a spectacular rise of water into the air. Offshore bars are small ridges of sand parallel to the shore that can also help dissipate the energy of breaking waves. Their presence is evident when incoming waves break one or more times before reaching the shore.

Beaches usually recover after each storm or storm season. The sand that moved offshore under the influence of the larger storm waves moves onshore under the influence of the smaller fair-weather waves. Offshore bars flatten or gradually move onshore, and the beach between the low- and high-tide lines tends to become steeper over time. The onshore distance depends on the size of the storm or severity of the storm season.

Beach grain size also affects the slope of the beach. High-permeability gravel for example absorbs much of the water that falls on it; hence, the return flow energy is much less than the incoming flow energy. In this case the forces pushing beach material up onto the beach are stronger than those moving material seaward in the backwash. The beach becomes steeper in response. On beaches of fine sand, the permeability of the material is much less than for gravel, so relatively little water from breaking waves is absorbed. The backwash of the wave has essentially the same volume of water as the upwash. This backwash volume tends to move sand in a seaward direction thereby flattening the beach.

### 6.1.7. Dunes

Beaches can be major sources of material for the land as well. Sand blown inland from the beach piles up as sand dunes, sometimes forming large dunes that move further inland. Some of these large dunes can be big enough to cover up trees and roads as the dunes move inland, as occurs on the US Pacific Coast near Florence, Oregon.

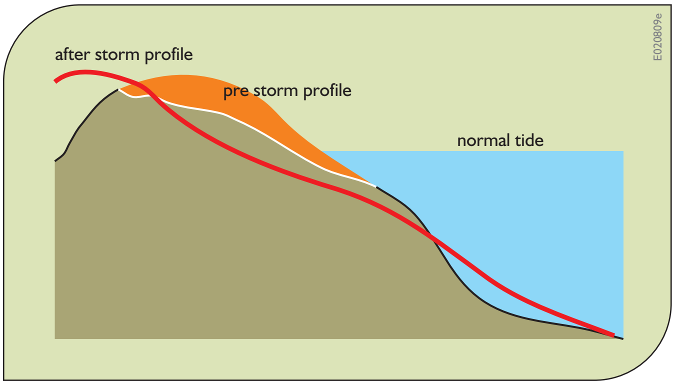



**Figure A.41.** Possible response of a beach to a major storm. Cross sections show steeper pre-storm profile, a possible shallower after-storm profile (denoted by the dark red line), and the recovering beach profile, showing some loss of its original steepness, sometime after the time of the storm.

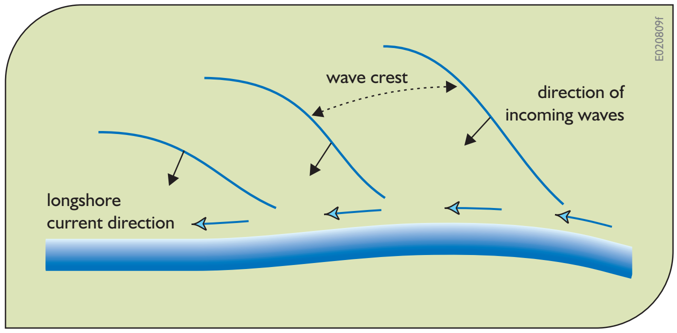



**Figure A.42.** The energy from breaking waves at the shoreline can cause longshore currents. These currents are responsible for transporting beach sand laterally along the beach.

### 6.1.8. Longshore Currents

Sand moves because of wind, waves and currents. The largest volume of sand is often carried by longshore currents. Such currents are created as waves strike the shoreline at an oblique angle, forcing some of the surf-zone water to move laterally along the shore, as illustrated in Figure A.42. Breaking waves suspend sand, and the currents move it. Sandy beaches are actually rivers of sand.

Beach material comes from the land as well as the seabed. Rivers supply sediment that eventually reaches the coast and ends up on beaches. Eroding cliffs along the coast are also sources of beach material. For most of the US East Coast, the continental shelf is an important source of sand, which is brought to the shore by the circular motion of the waves. Calcareous marine organisms are often contained in this continental shelf sand. In tropical waters the skeletal material of calcareous marine organisms can make up the entire beach, providing the sparkling white or pink colour characteristic of, for example, some Caribbean beaches.

## 6.2. Coasts under Stress

Many coastal areas are under pressure from a number of causes, mostly stemming from increasing human populations. Coastal regions tend to have substantially higher population densities than elsewhere. Today about half the US population lives 'near' (within easy driving distance of) the coast. In countries like Australia and Canada a much higher percentage of the population lives along a coast. Most of the world's major cities are located along coasts.

Historically the economy of coastal zones depended on fishing, boating or boat building. Today's populations are engaged in a much greater range of occupations that give them a higher standard of living and quality of life, even in increasingly expensive and crowded coastal regions.

Continued migrations of people to coastal regions for jobs, recreation, more moderate weather and a higher quality of life have increased the stress on coastal zone resources. Resorts along the Mediterranean coast, South Florida and the Caribbean islands, for example, have witnessed large increases in the tourism and recreation industries. Coastal areas along the western border of North America have seen substantial increases in residential development for people desiring the lifestyle associated with the coast.

Coastal zones are relatively scarce commodities. They are just narrow strips of land along a coast. The high demand for this limited real estate invariably results in relatively high land prices and in the costs of using various facilities located along the shore. In addition, many coastal lands are fragile. They erode easily. This puts a high priority on protecting and maintaining what is there.

## 6.3. Management Issues

Individuals like to live, work and play along coasts. This is evidenced by the continued higher proportion of population growth and economic development along

**Figure A.43.** Encroaching residential development along a coastal beach and estuary.

the coastal regions of just about every country in the world that has them. Most of the world's mega-cities (populations in excess of 10 million) are coastal. As suggested by Figure A.43, coastal areas provide attractive environments for residential as well as economic development in spite of their occasional natural hazards and changing landforms. Changing land cover and use that accompany population growth in coastal zones affect the geomorphological and ecological processes that take place along coastal shorelines.

Coastal zone management is the management of conflicts among multiple conflicting uses of natural and human resources in highly populated and economically, ecologically and often geopolitically important coastal zones. Coastal zone management focuses on the economic development and use of the environmental and ecological resources of coastal regions in ways that protect, and in some cases restore, them.

Coastal zone management historically focused on providing protection of shoreline properties against floods, and the design and construction of harbours and marinas (that is, the infrastructure associated with commercial and recreational boating). Today, coastal zone management involves much more than protection of land from being eroded by the sea and transportation on the sea. Issues such as water quality, dispersion of pollutants and the proper management of the complete coastal ecosystem have become important. Structural design is now only a small aspect of coastal management.

Current coastal zone management issues include reducing the risk of erosion from coastal storms, protecting bays, sloughs (channels) and sounds where rivers meet the sea, preserving and restoring beaches and ecological habitats, enhancing economic development and recreation opportunities along urban waterfronts, and controlling polluted runoff – the major source of pollution in coastal waters today.

The development of technically, economically and environmentally appropriate management plans requires an understanding of the complexity of coastal zone processes that probably no single person has mastered. It requires an understanding of the effects of the natural variability of wind and waves and temperatures as well as the effects of structural development, over-fishing, excessive pollution, habitat destruction and human population shifts. Inputs are needed from many individuals who have expertise in one or more aspects of coastal zone management and who are willing to work with others who have different expertise and experiences.

This necessary expertise and experience may have to include river basin management as well as coastal zone management. Some coasts are affected by what happens in upstream river basins. If those basins drain into coastal estuaries, it may be important to consider the effects of basin management on the coastal zones. For example, much of the sediment along some of North America's Pacific Ocean coasts is brought to those coasts by rivers that discharge their waters and sediment into the Pacific Ocean. Damming these rivers typically diminishes the sediment load, that is, the sand being transported to the coast. Much of the sediment that would otherwise end up on beaches is trapped behind the dams, even though these dams can be many hundreds of kilometres upstream of the beaches.

Consider, for example, the impact the High Aswan Dam has had on the Nile Delta. The policy of keeping the downstream flows from reaching flood stage has resulted in a shrinking of the delta at the mouth of the river as it discharges into the Mediterranean Sea. Similarly, reduced sediment loads in the Mississippi River, along with sea level rise, have contributed to the reduction of the delta of the Mississippi River on the Gulf Coast of the State of Louisiana.

### 6.3.1. Beaches or Buildings

For most of the world's beaches, if there were no people there would be no coastal zone management problems. As humans place obstacles such as houses, apartment buildings, highways and seawalls along a beach, the dynamic processes that maintain natural beaches are constrained. Shoreline erosion, and hence its retreat inland, is blocked. The beach jams up against these structural objects. This causes it to narrow, which in turn leads to a reduction in the supply of sand to adjacent beaches.

Since there is a worldwide tendency for all but very rocky coasts to erode, protecting and maintaining them is a high priority, particularly given the high economic value of coastal land. The economics of coastal zone protection and maintenance are complex. But is it seldom economically feasible to invest in shoreline protection unless there is very dense development or extensive tourism. Areas such as Miami Beach in southern Florida in the United States, the Gold Coast in Australia, Scheveningen in the Netherlands, Copacabana in Brazil and the Chicago and Toronto waterfronts on the Great Lakes are prime candidates for coastal protection. Agricultural areas and areas of single-family residential properties, on the other hand, are not. Yet protection decisions are not just based on economics. Decisions about what to protect are often political as well.

Consider, for example, the coastal management strategy of the Netherlands. Much of the Netherlands, being below sea level, depends on coastal protection. One single storm event in 1953 breached the dykes along the coast and drowned nearly 2,000 people. The loss of livestock and property was extensive. The result was a decision to provide protection from sea storms as severe as those expected once in every 10,000 years. This is a storm intensity having a probability of 0.00001 of occurring in any single year.

Alternative strategies for flood protection in the Netherlands include withdrawal from areas where further erosion may occur, selective erosion control, full erosion control, and expansion in the seaward direction by artificial beach nourishment where the coastal defences are considered to be weak. Full erosion control and perhaps expansion in threatened areas would appear to be almost a necessity for this densely populated and flood-vulnerable country. Full control involves placing 6–10 million $m^3$ of sand along the coast annually.

### 6.3.2. Groundwater

Freshwater in maritime coastal regions usually comes from two sources: rivers and lakes, and groundwater aquifers. When precipitation is sufficient to keep the groundwater levels higher than the surrounding sea level, the higher density saltwater does not mix with the freshwater. A floating freshwater lens will lie on top of the denser saltwater.

If withdrawals of freshwater from groundwater aquifers along coastal regions lower the freshwater level to below the seawater level, a flow of saltwater from the sea into the groundwater aquifer can occur. Since the flow rates of groundwater are very small, there is little mixing of the salt and fresh groundwater. The freshwater reservoirs of small island communities are therefore very susceptible to saltwater intrusion if population pressures along the coasts cause over-pumping of the aquifer for freshwater supplies. Similar damage can be caused by cutting away the dunes and thus lowering the water table, by increasing the amount of impervious surfaces that increases runoff and reduces aquifer recharge, and by dredging of rivers that can increase saltwater intrusion upstream.

### 6.3.3. Sea Level Rise

A potentially significant change in sea level may come from changes in the global climate. Global warming since the last glaciation has resulted in a sea level rise of 100 to 150 metres through melting of the polar ice caps and thermal expansion of the water in the ocean. This process continues at a rate estimated to be about 1 to 1.5 mm/yr. Any additional warming would increase this rate.

Global climate change models estimate the water level rise resulting from increases in greenhouse gases and hence increases in temperature, glacier ice melt and ocean thermal expansion. Such numerical models have produced a range of scenarios, and these predictions of water level increases are by no means precise. There are many uncertainties in the estimates of production of greenhouse gases. Probably the most uncertain is the extent to which gas emissions will be controlled, a matter that is unpredictable and largely political. The methods used to translate these uncertain atmospheric pollution figures first into global warming and then into water level rise also involve many assumptions that are uncertain. Nevertheless, the impacts of sea level rise are fairly clear:

- More severe storms (tornadoes and hurricanes) will occur more often.
- While storm surge will decrease a little because of the larger water depths, it will also increase because of the more severe storm activity.
- Offshore, the waves will be higher, because of more severe storms.
- Tides appear not to be significantly affected.
- Tidal prisms will increase, because the surface area of the bays and estuaries will increase.
- Waves breaking on shores and structures will be higher because larger depths all the way into shore will reduce bottom friction losses, as well as permitting large breaking waves to come closer into shore.

The result:

- Structures will be subjected to higher stress from the higher waves; factors of safety will decrease.
- Structure runup and overtopping will increase, adding to the risk of flooding and damage by overtopping. As an example, the Delta Project in the Netherlands raised all dykes in response to the 1953 storm surge, yet a 1-m sea level rise would reduce the present margin of safety by about 90% (Wind, 1987).
- Sandy shorelines will retreat.
- Barrier islands will roll back more rapidly and marshes behind the barrier islands will disappear.
- Deltas will not build out at the same rate; they may even retreat.
- Bluffs and cliffs will retreat more rapidly.
- Sediment transport rates will increase, possibly filling currently stable inlets and harbour entrance channels.
- Saltwater intrusion into groundwater tables will increase.
- Many wetland areas will be inundated and disappear.

How can everyone prepare? Flood protection, shore protection and navigation structures can be strengthened and raised to cope with the rise in water level. If all else fails, a properly executed retreat can be planned in which buildings are moved back from the shore or abandoned.

The magnitude of the problem can be huge. Population densities along the ocean shore are already high and rapidly increasing. Sustainable development of the coastal areas in the face of rising sea levels will be an important issue and challenge in this century. Moreover, although flood defences such as dykes can be raised with presently available technology, the risk to people and properties behind those dykes increases.

The main casualties will be the already limited wetland areas. Their development can keep up with slowly rising sea levels and move inland, but they may have problems adjusting to a more rapid rise in water levels. Also, since most of the properties behind the wetlands are usually dedicated to human activities and uses, it is unlikely that wetlands will be allowed to intrude into this valuable real estate. Damage will also occur to agricultural areas because of the additional saltwater intrusion.

Sustainable shoreline development through maintaining the existing shorelines by retro-fitting, retreat by moving infrastructure inland, and meeting the concerns for wetlands and agriculture, will require a complete restructuring of most current political and policy-making processes. They typically are not designed to deal with slowly developing impacts over very large areas.

### 6.3.4. Subsidence

Although subsidence can occur naturally, it is often caused by human activities, for example, pumping groundwater, petroleum and natural gas out of the ground. Subsidence of coastal landmass results in increased flooding. It exacerbates the effects of rising sea levels, since the relative sea level rise with respect to the land will be even greater. One of the best known examples of subsidence coupled with a rising sea level can be seen in the city of Venice, where increased pumping of both water and natural gas has caused an accelerated rate of subsidence. The result is more frequent flooding of an increasingly large portion of the city. As a result, the city and its mediaeval monuments are subjected more and more regularly to 'aqua alta' or high water.

### 6.3.5. Wastewater

Coastal waters have traditionally been used for wastewater disposal. Increased discharges of sewage and chemical effluents, particularly since the Second World War, have polluted many coastal waters. Although large oceans and lakes are often thought to have almost unlimited dilution capacity for wastewater, they do not. In many cases those dilution limits have already been reached, and many are showing serious overload and eutrophication problems. Even though some offending chemical dump sites have been cleaned up, the toxic chemical content of organisms in the aquatic ecosystem, including the fish for example, can remain at high levels for considerable times after cleanups are completed.

Cleanup of any coastal area is a complex undertaking, but even more so if it requires altering the pollutant discharge actions of many nations. There is still the prevailing philosophy of 'out of sight, out of mind', when it comes to the discharge of pollutants into maritime coastal waters.

Many maritime communities still discharge raw sewage, often into the near-shore zone. Even communities that have adequate sewage treatment prior to discharge often discharge sewage into coastal waters during heavy rainstorms due to combined sewer overflows (CSOs). With heavy rainfall, the combined sewage and stormdrain flows can exceed the hydraulic and treatment capacities of the sewage treatment plant, and the excess water (overflow) containing some raw sewage is discharged directly without treatment.

Beach pollution is less likely if the sewage outfall is piped and diffused into the coastal waters many kilometres offshore, as occurs at Sydney, Australia. Many outfalls are mere open channels that discharge storm and wastewaters relatively close to shore. The discharged pollutants, trapped by currents and wave action, can flow right along the coastline, leading to unacceptably high pollutant concentrations near the shore.

People have discharged raw sewage into the sea because it has been a relatively cheap way to dispose of it. If beach closures and the accompanying economic losses happen often enough, the added costs of wastewater treatment and disposal methods might be a less costly alternative. Tertiary wastewater treatment costs about ten times as much as dumping raw sewage. Incineration, if acceptable, costs even more.

Two recent examples of actions taken to clean up coastal zones, at considerable cost, have been those for reducing the discharge of pollutants into the estuary adjacent to the city of Boston on the northeastern Atlantic Coast of the United States (Figure A.44), and into the Tagus Estuary adjacent to the city of Lisbon in Portugal.

### 6.3.6. Other Pollutants

Runoff from farming can also be harmful to coastal waters, particularly bays and lagoons. Runoff that contains fertilizers used to promote the growth of agricultural products also promotes the growth of algae and aquatic weeds. Pesticides contain high levels of toxic substances, such as heavy metals. Even without the chemicals, runoff from farmlands can be undesirable. The manure from the high-density populations of cattle and pigs causes high levels of nitrates in both the groundwater and the surface waters. The runoff of fine sediment materials from soil erosion resulting from the conversion of forests to agricultural land has also caused problems for many maritime organisms. The death of coral reefs in many tropical countries can be attributed, at least in part, to the sediment that has entered the water column since the land was cleared for agriculture, as early as the eighteenth century in some cases.

Oil spills resulting from transportation and exploration of oil close to shore, such as the spill from the Exxon Valdez in 1989 (Figure A.45), have been the cause of well-known and costly disasters. The *Prestige* oil spill off the northwest coast of Spain in November 2002 involved twice as much oil as the *Exxon Valdez* spill. Areas along major shipping routes are particularly vulnerable, as witnessed by the spills of the *Torrey Canyon* in 1967, the *Amoco Cadiz* in 1978 and the *Erika* in 1999, all along the northwestern coast of France. Although such major incidents cause some government action, much larger volumes of oil are routinely released with virtually no restrictions each year into the oceans from ships and petroleum production platforms and refineries.

**Figure A.44.** Deer Island Sewage Treatment Plant in Boston Harbour treats and discharges an average daily flow of about 380 million gallons (1.43 million $m^3$) of wastewater. (With permission of Piping Resources Partnership of New England.)

Hazardous solid waste is another major source of pollution of coastal waters. Although ocean dumping legislation severely limits the dumping of solid waste in many countries, the oceans continue to be the recipient of all sorts of solid hazardous wastes, including hospital waste, contaminated dredge spoil and nuclear material.

### 6.3.7. Mining of Beach Materials

Mining of sand and other beach materials is one of the major causes of beach erosion in some beaches. Sand mining for construction and other uses occurs on many scales, but the accumulated impact can be substantial along some coasts.

## 6.4. Management Measures

Coastal zone management is the management of the land and its uses along the coast. Thus it is the management of conflicts. It requires legislation and enforcement of measures to protect and enhance the sustainable use of lands along shorelines. Coastal zone managers need scientific, technical and political skills, skills based on an understanding of geology and morphology, biology, ecology, law, planning and engineering among other disciplines. Although coastal zone management is interdisciplinary, it is usually the engineers who are asked to make and implement crucial technical decisions. To do this well, they need to be properly informed, and work with specialists from many other disciplines.

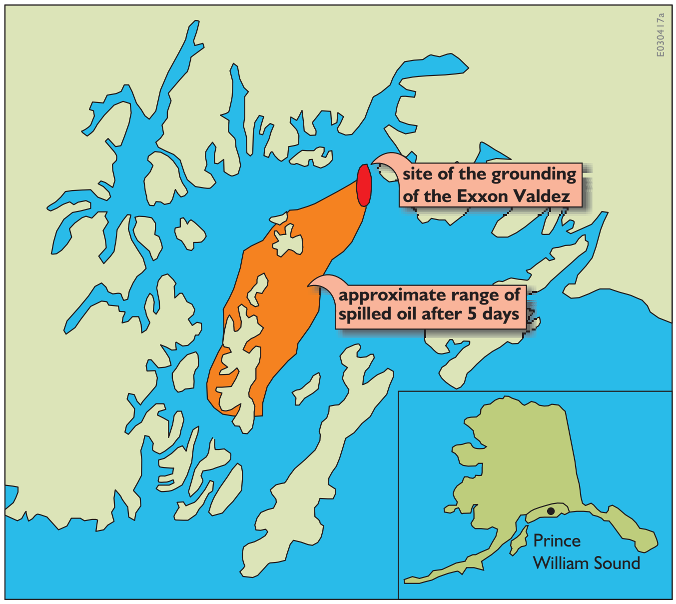



**Figure A.45.** Site of Exxon Valdez oil spill in the US State of Alaska, 1989.

Management options include structural as well as non-structural measures such as zoning, regulation enforcement, public awareness activities, negotiation and consultation. These tools need to be part of an appropriate adaptive decision-making process.

### 6.4.1. 'Conforming Use'

Coastal zones have traditionally supported many activities, all of which compete for limited space and may or may not conflict with each other. Some of the more important uses of coastal lands include tourism, residential, recreational, industrial and commercial activities, agriculture, transportation, waste disposal, aquaculture, fishing, nature reserves, and military and security functions. (Not all of these may be present along any given coast.)

Given the high demand for limited space along a coast, one approach to coastal zone resource management is to limit development to only those 'conforming uses' that need to be along the coast. Examples could include swimming beaches, fishing ports and marinas. Amusement parks, casinos, theatres and car parking areas that are often found along a coast can be located more inland from the beach. Development (building) permits for beach property could be limited to those projects that need to be there.

Implementing a restrictive zoning policy is not simple. For instance, developments that attract people to an area, such as harbours or marinas, are normally required by law to supply sufficient infrastructure to support their operation. Does a car park suddenly become a conforming use, if it is needed in support of a conforming use? Similarly, money generators such as casinos are usually permitted to locate along the coast regardless of conforming use. The installation of adequate public transportation from inland casinos, hotels and parking lots to coastal shorelines could alleviate some of this political as well as economic pressure for non-conforming uses right along shorelines.

In many coastal regions, abandoned older facilities are being taken over by interests catering to tourism. Facilities such as factories, ports and railway lines are being converted to more conforming uses. Warehouses and loading terminals are being transformed into apartment buildings, abandoned railway lines are serving as biking and hiking trails, and small commercial ports are becoming marinas and condominiums. These transformations are taking advantage of unique opportunities, but they also present unique challenges.

In environmentally sensitive regions, tourism can be a positive influence. Proper coastal management is, in part, aimed at enhancing the intrinsic value of the area for tourism. The definition of intrinsic value of the coast is also changing. Highways, parking lots, hotels and fast-food restaurants are typical features of many tourist areas, but an increasing number of tourists demand other aspects, especially along coastlines. They seek bicycle paths, dunes, wetlands, clean water, birds and fish. Increasing numbers of tourists are more interested in nature and prefer physical activities such as hiking, biking, birding, boating and fishing to simply sitting on the beach soaking up the sun and a few six-packs. The coastal environment they want is much more natural than the traditional amusement park type of environment dominated by concrete and French fries. (Obviously, many tourists like both the concrete and the sun and sand.)

Whatever tourists seek, tourism development is clearly motivated by economics. Enhanced recreational (as well as environmental and ecosystem) planning must be done within an economic framework that sustains the continued coastal zone development, maintenance, protection and restoration activities.

### 6.4.2. Structures

Many types of structures are built along coastlines to meet the needs of people who visit or live there. The focus here is only on those that are built to protect property from damage by waves and currents along the coasts.

There are three available responses to local shoreline erosion problems:

- hard stabilization, such as seawalls (Figure A.46), offshore breakwaters (Figure A.47) and groynes (Figure A.48)
- soft stabilization, such as beach sand replenishment
- relocation of threatened structures.

Each response has its advantages and disadvantages. Hard stabilization alternatives involve armouring the shoreline with structures designed to hold it in place. This may be the best way to save buildings, but not beaches. Removing buildings or relocating them further inland is the best way to save beaches, but obviously not the buildings in their current locations. Hard stabilization combined with beach replenishment may provide necessary protection for both buildings and beaches, but it will be necessary to continually add sand to the beach over time.

### 6.4.3. Artificial Beach Nourishment

Artificial nourishment is an alternative to structural measures for protection. It has the least impact on adjacent properties and the environment. Instead of harming the surroundings, a beach fill can benefit adjacent eroding properties. Apart from its cost, the diffusion and advection of a beach fill only presents a problem when water depth needs to be maintained at the adjacent properties, such as in navigation channels, or when the sand added to the system threatens valuable habitat.

Although this approach has been used for many years, the technology is still very much intuitive. Artificial beach nourishment is subject to the same erosion that caused the need for replenishment in the first place. The design of such programmes must therefore be concerned not only with how much sand to place, but also with how often it needs to be replenished. Artificial nourishment in most areas then becomes a beach maintenance programme based on annual cost–benefit estimates. If the site erodes more rapidly as a result of offshore conditions, such as a locally steeper shoreline or a convergence of wave energy, the artificially placed fill will also be subjected to the same conditions and will not perform well.

Since a major objective of most artificial nourishment schemes is to provide protection as well as additional recreational beach space, most schemes consider beach fills in combination with shore face nourishment. Beach fill normally requires rehandling of the sand so that it can be placed by pipeline dredge and perhaps be reshaped by earthmoving equipment. The onshore sand is usually placed with a steep seaward slope. The wave action on such a fill will shape the most seaward part of the fill mass into a beach profile. During this adjustment period – and at any later time, when other beach material further landward is redistributed – fine grain sizes will be winnowed out of the mass of sand and lost to deep water, until the grain size distribution of the remaining sand mass is similar to the native distribution. Once the fill has been re-adjusted by the waves to form a beach profile, a steep scarp may have formed at the top of the beach.

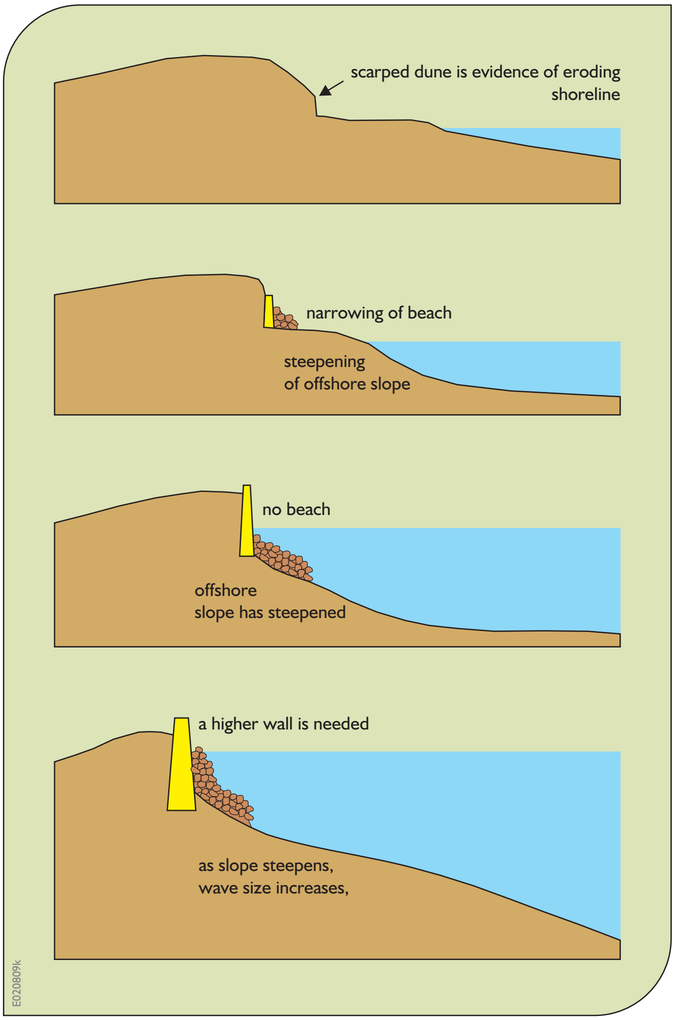



**Figure A.46. Seawalls designed to protect property. Larger and stronger seawalls are needed over time as the beach disappears in front of them.**

The combination of artificial nourishment with structures such as groynes or offshore breakwaters will help contain the fill material. Structures also provide an opportunity to use beach fills in areas that would never be stable with artificial nourishment alone. Examples are Hilton Head on the US Atlantic Coast (Bodge et al., 1993) and Norderney on Europe's North Sea Coast (Kunz, 1993).

Biologically, a beach is relatively unproductive. Benthic communities that are covered by a beach fill seem to re-establish relatively quickly after nourishment. The surrounding ecosystem, however, can be affected and thus

**Figure A.47.** Eastern Scheldt Storm Surge Barrier along the North Sea coast of the Netherlands, a seawall eight kilometres long. Its sixty-two openings are closed off with sliding barriers at high tide.

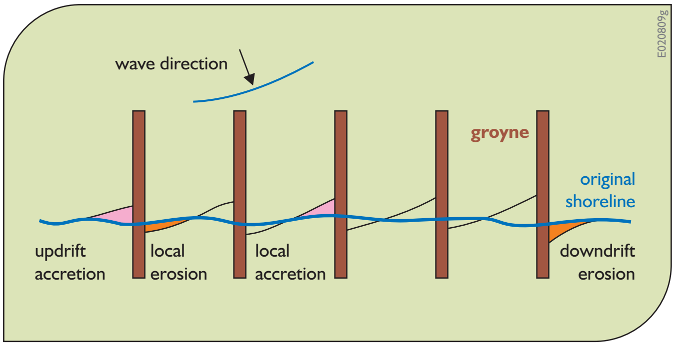



**Figure A.48.** Groynes are often used to reduce shoreline erosion. They change the longshore sediment transport rates and result in accretion updrift of the groynes within the groyne field and downdrift erosion.

should be carefully considered and monitored (CUR, 1997; NRC, 1995; Simm, 1996).

## 7. Conclusion

This concludes an overview of the many components, functions and processes that take place in watersheds and river basins, wetlands, estuaries and along coasts, and the ways in which these processes can be affected by land and water management policies and practices. This appendix provides an introductory background for those who have not had separate courses in these subjects and who are now being introduced, through this book, to the art of building and solving water resources planning and management models. The purpose of these models is to predict how, and to what extent, these system components, functions and processes are affected by alternative ways of managing and using these land and water resources.

## 8. References

BODGE, K.R.; OLSON, E.J. and CREED, C.G. 1993. Performance of beach nourishment at Hilton Head Island, S.C.: beach nourishment engineering and management considerations. In: *Proceedings of the Eighth Symposium on Coastal and Ocean Management (Coastal Zone '93), New Orleans, La., 19–23 July 1993*. New York, ASCE, pp. 668–76.

BOWDEN, K.F. 1967. Circulation and diffusion. In: G.H. Lauff (ed.), *Estuaries*. Washington, D.C., American Association for the Advancement of Science (Publ. No. 83). pp. 15–36.

BRICE, J.C. and BLODGETT, J.C. 1978. *Countermeasures for hydraulic problems at bridges. Volume 1: Analysis and assessment. Volume 2: Case histories for sites 1–283*. Federal Highway Administration report (FHWA-RD-78-162 and FHWA-RD-78-163).

CORMACK, D. and NICHOLS. 1977. The concentrations of oil in seawater resulting from natural and chemically induced dispersion of oil slicks. In: J.A. Nichols (ed.), *Proceedings of the Oil Spill Conference (Prevention, Behaviour Control, Clean-up), New Orleans, March 1977*. Washington, D.C., American Petroleum Institution (Publication 4284). pp. 381–5.

COVICH, A.P. 1993. Water and ecosystems. In: P.H. Gleick (ed.), *Water in crisis: a guide to the world's freshwater resources*. Oxford, UK, Oxford University Press.

COWARDIN, L.W.; CARTER, V.; GOLET, F.C. and LAROE, E.T. 1979. *Classification of wetlands and deep water habitats of the United States*. Washington, D.C., US Fish and Wildlife Service.

CULBERTSON, D.M.; YOUNG, L.E. and BRICE, J.C. 1967. *Scour and fill in alluvial channels*. Washington, D.C., USGS (Open-File Report).

CUR. 1997. *Beach nourishment and shore parallel structures*. Gouda, Netherlands, CUR, (Rep 97-2).

DEAN, R.G. and DALRYMPLE, R.A. 1984. *Water wave mechanics for engineers and scientists*. Englewood Cliffs, N.J., Prentice Hall.

DINGEMANS, M.W. 1997. *Water wave propagation over uneven bottoms. Advanced series on ocean engineering*. Singapore, World Scientific Publishing. Vol. 13.

FORMAN, R.T.T. and GODRON, M. 1986. *Landscape ecology*. New York, Wiley.

HARDY, J.T.; CRECELIUS, E.A.; ANTRIM, L.D.; KIESSER, S.L. and BROADHURST, V.L. 1990. Aquatic surface microlayer contamination in Chesapeake Bay. *Marine Chemistry*, No. 28, pp. 333–51.

HORIKAWA, K. 1978. *Coastal engineering*. Tokyo, University of Tokyo Press.

HORTON, R.E. 1945. Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin*, No. 56, pp. 275–370.

HUTCHINGS, P.A. and COLLETT, L.C. 1977. Guidelines for the protection and management of estuaries and estuarine wetlands. Toowong, Queensland, Australian Marine Sciences Association.

IPPEN, A. 1966. *Estuary and coastline hydrodynamics*. New York, McGraw-Hill.

KINSMAN, B. 1965. *Wind waves*. Englewood Cliffs, N.J., Prentice Hall.

KUNZ, H. 1993. Sand losses from artificially nourished beach stabilized by groynes, beach nourishment engineering and management considerations. In: *Proceedings of the Eighth Symposium on Coastal and Ocean Management (Coastal Zone '93), New Orleans, La., 19–23 July 1993*. New York, ASCE, pp. 668–76.

LAGASSE, P.F.; SCHALL, J.D.; JOHNSON, F.; RICHARDSON, E.V.; RICHARDSON, J.R. and CHANGE, F. 1991. *Stream stability at highway structures*. McLean, Va., HEC-20 (publ. No. FHWA-IP-90-014).

LANE, E.W. 1955. The importance of fluvial morphology in hydraulic engineering. *Proceedings of the American Society of Civil Engineers*, No. 81, Paper 745, pp. 1–17.

LEOPOLD, L.B. 1994. *A view of the river*. Cambridge, Mass., Harvard University Press.

LEOPOLD, L.G. and MADDOCK, T. Jr. 1953 *The hydraulic geometry of stream channels and some physiographic implications*. Washington, D.C., USGS (Professional Paper 252).

LEOPOLD, L.B. and WOLMAN, M.G. 1957. River channel patterns: braided, meandering, and straight. In: *Physiographic and hydraulic studies of rivers*. Washington, D.C., US Government Printing Office (Professional paper US Geological Survey, Vol. 282B), pp. 39–85.

LEOPOLD, L.B.; WOLMAN, M.G. and MILLER, J.P. 1964. *Fluvial processes in geomorphology*. San Francisco, Calif., W.H. Freeman.

LEOPOLD, L.B.; WOLMAN, M.G. and MILLER, J.P. 1992. *Fluvial processes in geomorphology*. Mineola, N.Y., Dover Publications.

LIU, P.L.F. and LOSADA, I.J. 2002. Wave propagation modeling in coastal engineering. *Journal of Hydraulic Research*, Vol. 40, No. 3, pp. 229–40.

MADSEN, K.N.; NILSSON, P. and SUNDBACK, K. 1993. The influence of benthic microalgae on the stability of a subtidal sediment. *Journal of Experimental Marine Biology and Ecology*, No. 170, pp. 159–77.

MASER, C. and SEDELL, J.R. 1994. From the forest to the Sea: the ecology of wood in streams, rivers, estuaries and oceans, St. Lucie Press, Delray Beach, Florida, p 200.

MCDOWELL, D.M. and O'CONNOR, B.A. 1977. *Hydraulic behaviour of estuaries*. London, MacMillan.

MEADOWS, P.S.; TAIT, J. and HUSSAIN, S.A. 1990. Effects of estuarine infauna on sediment stability and particle sedimentation. *Hydrobiologia*, No. 190, pp. 263–6.

MEI, C.C. and LIU, P.L.F. 1993. Surface waves and coastal dynamics. *Annual Review of Fluid Mechanics*, No. 25, pp. 215–40.

MINSHALL, G.W.; CUMMINS, K.W.; PETERSEN, R.C.; CUSHING, C.E.; BRUNS, D.A.; SEDELL, J.R. and VANNOTE, R.L. 1985. Developments in stream ecosystem theory. *Canadian Journal of Fisheries and Aquatic Sciences*, No. 42, pp. 1045–55.

NRC (National Research Council). 1995. Beach nourishment and protection. Washington, D.C., National Academy Press.

POFF, N.L.; ALLAN, J.D.; BAIN, M.B.; KARR, J.R.; PRESTEGAARD, K.L.; RICHTER, B.D.; SPARKS, R.E. and STROMBERG, J.C. 1997. The natural flow regime: a paradigm for river conservation and restoration. *BioScience*, No. 47, 1997, pp. 769–84.

PRITCHARD, D.W. 1952. Estuarine hydrography. *Advances in Geophysics*, No. 1, pp. 243–80.

RICHARDSON, E.V.; SIMONS, D.B. and JULIEN, P.Y. 1990. *Highways in the river environment*. Washington, D.C., FHWA, (revision of 1975 publication by E.V. Richardson, D.B. Simons, S. Karki, K. Mahmood and M.A. Stevens; Pub. No. FHWA-HI-90-016).

RHOADS, D.C. 1974. Organism–sediment relations on the muddy sea floor. *Oceanography Marine Biology Annual Review*, No. 12, pp. 263–300.

RHOADS, D.C. and BOYER, L.F. 1982. The effects of marine benthos on physical properties of sediments. In: P.L. McCall and M.J.S. Tevesz, (eds.), *Animal–sediment relations: the biogenic alteration of sediments*. New York, Plenum, pp. 3–52.

ROSGEN, D. 1996. *Applied river morphology*. Pagosa Springs, Colo., Wildland Hydrology.

SARPKAYA, T. and ISAACSON, M. 1081. *Mechanics of wave forces on offshore structures*. New York, Van Nostrand-Reinhold.

SCHUMM, S.A. (ed.). 1972. *River morphology: benchmark papers in geoplogy*. Stroudsburg, Pa., Dowden, Hutchinson and Ross.

SCHUMM, S.A. 1977. *The fluvial system*. New York, Wiley.

SIMM, J. 1996 *Beach management manual*. London, Construction Industry Research and Information Association (CIRIA) (Rep. 154).

SPARKS, R. 1995. Need for ecosystem management of large rivers and their flooodplains. *BioScience*, Vol. 45, No. 3, p. 170.

SZEKIELDA, K.H.; KUPFERMAN, S.I.; KLEMAS, T. and POLIS, D.F. 1972. Element enrichment in organic films and foams associated with aquatic frontal systems. *Journal of Geophysical Research*, Vol. 77, pp. 5278–84.

TRIMBLE, S.W. and CROSSON, P. 2000. U.S. soil erosion rates: myth and reality. *Science*, Vol. 289, pp. 248–50.

USDA (United States Department of Agriculture). 1998. *Stream corridor restoration: principles, processes and practices*. Washington, D.C. Interagency Stream Restoration Working Group; (See also Federal Interagency Stream Restoration Working Group, (FISRWG), *Stream corridor restoration: principles, processes, and practices*, GPO Item No. 0120-A; SuDocs No. A 57.6/2:EN3/PT.653. ISBN-0-934213-59-3 2001.)

USEPA (United States Environmental United States Environmental Protection Agency) 2001. www.usepa.gov (accessed 12 November 2004).

USACE (US Army Corps of Engineers). 1969. *Missouri river channel regime studies*. Omaha, Nebr., USACE (MRD Sedimentation Series No. 13.B).

WARD, A.D. and ELLIOT, W.J. 1995. *Environmental hydrology*. Boca Raton, Fla., CRC Lewis 1995.

WIND, H.G. 1987. *Impact of sea level rise on society*. Rotterdam, Balkema.

## Additional References (Further Reading)

ABBOTT, M.B. and Price, W.A. 1994. *Coastal, estuarial and harbour engineers' reference book*. London and New York, E.&F.N. Spon.

ABERNETHY, B. and RUTHERFORD, I.D. 1998. Where along a river's length will vegetation most effectively stabilize stream banks? *Geomorphology*, Vol. 23, pp. 55–75.

ABT, S.R.; CLARY, W.P. and THORNTON, C.I. 1994. Sediment deposition and entrapment in vegetated streambeds. *Journal of Irrigation and Drainage Engineering*, Vol. 120, No. 6, pp. 1098–111.

ACKERS, P. and CHARLTON. 1970. Meandering geometry arising from varying flows. *Journal of Hydrology*, Vol. 11, No. 3, pp. 230–52.

ACKERS, P. and WHITE, W.R. 1973. Sediment transport: new approach and analysis. *Journal of the Hydraulics Division, ASCE 99, No. HY11* (American Society of Civil Engineers Proc. Paper 10167), pp. 2041–60.

ADAMS, W.J.; KIMERLE, R.A. and BARNETT, J.W. Jr. 1992. Sediment quality and aquatic life assessment. *Environmental Science and Technology*, Vol. 26, No. 10, pp. 1864–75.

ALDRIDGE, B.N. and GARRETT, J.M. 1973. *Roughness coefficients for stream channels in Arizona*. Washington, D.C., US Government Printing Office (US Geological Survey, open-file report).

ALLAN, J.D. 1995. *Stream ecology: structure and function of running waters*. New York, Chapman and Hall.

BAYLEY, P.B. 1995. Understanding large river-floodplain ecosystems. *BioScience*, Vol. 45, No. 3, p. 154.

BESCHTA, R.L.; PLATTS, W.S.; KAUFFMAN, J.B. and HILL, M.T. 1995. Artificial stream restoration: money well spent or an expensive failure? In: *Proceedings on environmental restoration:, UNCOWR 1994 annual meeting*. Southern Illinois University, Carbondale, Ill. pp. 76–104.

BIEDENHARN, D.S.; ELLIOTT, C.M. and WATSON, C.C. 1997. *The WES stream investigation and streambank stabilization handbook*. Vicksburg, Miss., prepared for US Environmental Protection Agency by US Army Corps of Engineers, Waterways Experiment Station.

BIRD, E.C.F. 1984. *Coasts*, third edn. Oxford, UK, Basil Blackwell.

BIRD, E.C.F. 1985. *Coastline changes: a global review*. New York, Wiley.

BIRD, E.C.F. 1993. *Submerging coasts: the effects of a rising sea level on coastal environments*. New York, Wiley.

BISHOP, P.L. 1983. *Marine pollution and its control*. New York, McGraw-Hill.

BLENCH, T. 1957. *Regime behaviour of canals and rivers*. London, Butterworths Scientific.

BOOTH, D. and JACKSON, C. 1997. Urbanization of aquatic systems: degradation thresholds, stormwater detection and the limits of mitigation. *Journal of the American Water Resources Association*, Vol. 33, No. 5, pp. 1077–89.

BREUSERS, H.N.C. and RAUDKIVI, A.J. 1991. *Scouring*. Rotterdam, Balkema.

BRICE, J.C. 1984. Planform properties of meandering rivers. In: C.M. Elliott (ed.), *River meandering*, pp. 1–15. Proceedings of the 1983 Conference on Rivers. New Orleans, ASCE.

BROWN, A.G. 1997. Biogeomorphology and diversity in multiple-channel river systems. *Global Ecology and Biogeography Letters*, Vol. 6, Floodplain forests special issue, pp. 179–85.

BROWN, A.G.; HARPER, D. and PETERKEN, G.F. 1997. European floodplain forests: structure, functioning and management. *Global Ecology and Biogeography Letters*, Vol. 6, Floodplain forests special issue, pp. 169–78.

BROWN, C.B. 1950. Sediment transportation. In: H. Rouse (ed.), *Engineering hydraulics*, pp. 769–858. New York, Wiley.

BROOKES, A. 1990. Restoration and enhancement of engineered river channels: some European experiences. *Regulated Rivers: Research and Management*, Vol. 5, No. 1, pp. 45–56.

BROOKES, A. and SHIELDS, F.D. Jr. (eds.). 1995. *River channel restoration: guiding principles for sustainable projects*. Chichester, UK, Wiley.

CHESAPEAKE BAY PROGRAM OFFICE. 2001. www.chesapeakebay.net/wqcmodelling.htm (accessed 12 January 2004).

COLBY, B.R. 1964. *Discharge of sands and mean-velocity in sand-bed streams*. Washington, D.C., USGS (Professional Paper 462-A).

CUMMINS, K.W. 1974. The structure and function of stream ecosystems. *BioScience*, Vol. 24, pp. 631–41.

DALRYMPLE, R.A.; KIRBY, R.T. and HWANG, P.A. 1984. Wave diffraction due to area of energy dissipation. *Journal of Waterway, Port, Coastal, and Ocean Engineering, ASCE*, Vol. 110 , No. 1, pp. 67–79.

DEAN, R.G. 1983. Principles of beach nourishment. In: P. Komar (ed.), *Handbook of coastal processes and erosion*. Boca Raton, Fla., CRC Press, pp. 217–32.

DELFT HYDRAULICS; DHI. 1996. *Qualitative assessment of possible morphological impacts of FAP implementation*. Delft, the Netherlands, Water Resources Planning Organization, Government of Bangladesh (River Survey Project Special Report No. 5).

DELFTHYDRAULICS; DHI and EGIS. 1996. *Spatial representation and analysis of hydraulic and morphological data*. Delft, the Netherlands, Water Resources Planning Organization, Government of Bangladesh (River Survey Project Special Report No. 17).

DE VRIEND, H.J. 2003. On the prediction of aggregated-scale coastal evolution. *Journal of Coastal Research*, Vol. 19, No. 4, pp. 757–9.

DOWNS, P.W. 1995. River channel classification for channel management purposes. In: A. Gurnell and G. Petts (eds.), *Changing river channels*. New York, Wiley, pp. 347–65.

DRUERY, B.M.; DYSON, A.R. and GREENTREE, G.S. 1983. *Fundamentals of tidal propagation in estuaries*. Proceedings of the Sixth Aust. Conference on Coastal and Ocean Engineering. Barton, Engineers Australia.

DUNNE, T. and LEOPOLD, L.B. 1978. *Water in environmental planning*. San Francisco, Calif., W.H. Freeman.

DYER, K.R. 1989. Estuarine flow interaction with topography, lateral and longitudinal effects. In: B.J. Neilson, A. Kuo and J. Brubaker (eds.), *Estuarine circulation*. Clifton, N.J. Humana. pp. 39–59.

DYER, K.R. 1990. The rich diversity of estuaries. *Estuaries*, No. 13, pp. 504–5.

GIS and DELFT HYDRAULICS. 1997. *Morphological dynamics of the Brahmaputra-Jamuna River*. Delft, Environmental and GIS support project for water sector planning (EGIS).

EINSTEIN, H.A. 1950. *The bed-load function for sediment transportation in open-channel flow*. Washington, D.C., USDA-SCS (Technical Bulletin 1026).

ENGELUND, 1974. F. Flow and bed topography in channel bends. *Journal of the Hydraulics Division,* ASCE, Vol. 100, No. HY11, pp. 1631–48.

EXNER, F.M. 1925. *Über die Wechselwirkung zwischen Wasser und Geschiebe in Flüssen*, pt.IIa, Bd.134. Wien, Sitzungsberichte Akademie der Wissenschaften.

FERGUSON, B.K. 1991. Urban stream reclamation. *Journal of Soil and Water Conservation*, Vol. 46, No. 5, pp. 324–8.

FORMAN, R.T.T. 1995. *Land mosaics: the ecology of landscapes and regions*. Cambridge, UK, Cambridge University Press.

FRISSELL, C.A.; LISS, W.J.; WARREN, C.E. and HURLEY, M.D. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environmental Management*, Vol. 10, No. 2, pp. 199–214.

GERMANOSKI, G. and SCHUMM, S.A. 1993. Changes in braided river morphology resulting from aggradation and degradation. *Journal of Geology*, Vol. 101, pp. 451–66.

GOODWIN, C.N.; HAWKINS, C.P. and KERSHNER, J.L. 1997. Riparian restoration in the western United States: overview and perspective. *Restoration Ecology*, Vol. 5, No. 4S, pp. 4–14.

GORDON, N.D.; MCMAHON, T.A. and FINLAYSON, B.L. 1992. *Stream hydrology: an introduction for ecologists*. Chichester, UK, Wiley.

GRAF, W.H. 1971. *Hydraulics of sediment transport*. New York, McGraw-Hill, New York.

GRAF, W.L. (ed.). 1987. *Geomorphic systems of North America: Centennial special volume 2*. Boulder, Colo., Geological Society of America.

GREGORY, K.J. 1992. Vegetation and river channel process interactions. In: P.J. Boon, P. Calow and G.E. Petts (eds.), *River conservation and management*. Chichester, UK, Wiley, pp. 255–69.

GREGORY, K.J. and GURNELL, A.M. 1988. Vegetation and river channel form and process. In: H.A. Viles (ed.), *Biogeomorphology*. Oxford, Basil Blackwell, pp. 11–42.

GREGORY, S.V.; SWANSON, F.J.; MCKEE, W.A. and CUMMINS, K.W. 1991. An ecosystem perspective of riparian zones: focus on links between land and water. *BioScience*, Vol. 41, No. 8, pp. 540–51.

GURNELL, A.M. 1997. The hydrological and geomorphological significance of forested floodplains. *Global Ecology and Biogeography Letters*, Vol. 6, Floodplain forests special issue, pp. 219–29.

GURNELL, A.M. and GREGORY, K.J. 1995. Interactions between semi-natural vegetation and hydrogeomorphological processes. *Geomorphology*, Vol. 13, pp. 49–69.

HOFFMANS, G.J.C.M. and VERHEIJ, H.J. 1997. *Scour manual*. Rotterdam, Balkema.

HUPP, C.R. and OSTERKAMP, W.R. 1996. Riparian vegetation and fluvial geomorphic processes. *Geomorphology*, Vol. 14, pp. 277–95.

INGRAM, J.J. 1988. *Total sediment load measurement using point source suspended sediment data*. Vicksburg, Miss., USACE (Hydraulics Laboratory Technical Report HL-88-28, Waterways Experiment Station).

INTERAGENCY ECOSYSTEM MANAGEMENT TASK FORCE. 1995. *The ecosystem approach: healthy ecosystems and sustainable economies*; Vol. I, Overview. Washington, D.C., Council on Environmental Quality.

JAGERS, H.R.A. 1996. *Behaviour oriented modelling of braided rivers*. Delft, the Netherlands, Delft Hydraulics and University of Twente (research report), January.

JAGERS, H.R.A. 1997. On prediction models for planform changes of braided rivers. In: *Abstracts 6th Int. Conf. Fluvial Sedimentology, University of Cape Town, Cape Town, 22–6 September,* p. 93.

JAGERS, H.R.A. 2000. Draft of Ph.D. thesis. Netherlands, University of Twente.

JANSEN, P.P., VAN BENDEGOM, L.; VAN DEN BERG, J.; DE VRIES, M. and ZANEN, A. 1979. *Principles of river engineering: the non-tidal alluvial river*. London, Pitman.

KALKWIJK, J.P.T. and DE VRIEND, H.J. 1980. Computation of the flow in shallow river bends. *Journal of Hydraulic Resesearch,* IAHR, Vol. 18, No. 4, pp. 327–42.

KAMPHUIS, J.W. 2000. *Introduction to coastal engineering and management*. Singapore, World Scientific Publishing.

KARAMBAS, T.V. and KOUTITAS, C. 1992. A breaking wave propagation model based on the Boussinesq equations. *Coastal Engineering*, No. 18, pp. 1–19.

KAWAMURA, T. 1998. Numerical simulation of 3D turbulent free-surface flows. *Hørsholm,* Denmark, International Research Center for Computational Hydrodynamics (ICCH).

KELLERHALS, R.; CHURCH, M. and BRAY, D.I. 1976. Classification and analysis of river processes. *Journal of the Hydraulics Division,* ASCE, Vol. 102, No. HY7, pp. 813–29.

KENNEDY, A.B.; CHEN, Q.; KIRBY, J.T. and DALRYMPLE, R.A. 2000. Boussinesq modelling of wave transformation, breaking and runup, I: 1D. *Journal of Waterway, Port, Coastal, and Ocean Engineering, ASCE*, Vol. 126, No. 1, pp. 39–48.

KENTULA, M.E.; BROOKS, R.E.; GWIN, S.E.; HOLLAND, C.C.; SHERMAN, A.D. and SINFEOS, J.C. 1992. *An approach to improving decision-making in wetland restoration and creation*. Washington, D.C., Island Press.

KERCHNER, J.L. 1997. Setting riparian/aquatic restoration objectives within a watershed context. *Restoration Ecology*, Vol. 5, No. 45, pp. 459–71.

KERN, K. 1992. Restoration of lowland rivers: the German experience. In: P.A. Carling and G.E. Petts (eds.), *Lowland floodplain rivers: geomorphological perspectives*. Chichester, UK, Wiley, pp. 279–97.

KIRKBY, M. 1990. The landscape viewed through models. *Zeitschrift fur Geomorphologie, Suppl.-Bd.*, No. 79, pp. 63–81.

KIRKBY, M.J. (ed.). 1994. *Process models and theoretical geomorphology*. New York, Wiley.

KLAASSEN, G.J. and MASSELINK, G. 1992. Planform changes of a braided river with fine sand as bed and bank material. In: *Proceedings of the Fifth International Symposium on River Sedimentation*, pp. 459–71, Karlsruhe.

KLAASSEN, G.J.; MOSSELMAN, E. and BRÜHL, H. 1993. On the prediction of planform changes in braided sand-bed rivers. In: S.S.Y. Wang (ed.), *Advances in Hydro-Science and Engineering*. Oxford, Mass, University of Mississippi, pp. 134–46.

KLIJN, F. 1997. *A hierarchical approach to ecosystems and its implications for ecological land classification: with examples of ecoregions, ecodistricts and ecoseries of the Netherlands*. Ph.D. thesis, Leiden University.

KLIMAS, C.V. 1987. River regulation effects on floodplain hydrology and ecology. In: D.D. Hook (ed.), *The ecology and management of wetlands*, Chapter 4. London, Croom Helm.

KNIGHTON, A.D. and NANSON, G.C. 1993. Anastomosis and the continuum of channel pattern. *Earth Surface Processes and Landforms, BGRG*, Vol. 18, No. 7, pp. 613–25.

KNOPF, F.L.; JOHNSON, R.R.; RICH, T.; SAMSON, F.B. and SZARO, R.C. 1988. Conservation of riparian systems in the United States. *Wilson Bulletin*, No. 100, pp. 272–84.

LARGE, A.R.G. and PETTS, G.E. 1994. Rehabilitation of river margins. In: P. Calow and G. Petts (eds.), *The rivers handbook*, pp. 401–18. Oxford, UK, Blackwell Scientific.

LEWIS, R.E. 1997. *Dispersion in estuaries and coastal waters*. Chichester, UK, Wiley.

LEWIS, R.R. III. 1989. *Wetland restoration/creation/enhancement terminology: suggestions for standardization*. Washington, D.C., US Environmental Protection Agency (Wetland Creation and Restoration: The Status of the Science, Vol. II. EPA 600/3/89/038B).

LI, R.M. and SHEN, H.W. 1973. Effect of tall vegetations on flow and sediment. *Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers*, Vol. 99, No. HY5, pp. 793–814.

LIN, P. and LIU, P.L.F. 1998. A numerical study of breaking waves in the surf zone. *Journal of Fluid Mechanics*, No. 359, pp. 239–64.

LORENZ, E. 1963. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, Vol. 20, pp. 130–41.

MALANSON, G.P. 1993. *Riparian landscapes*. Cambridge, UK, Cambridge University Press.

MANLEY, P.A.; HAZELHURST, S.; MAGARY, F. and HAWK, K.S. 1995. *Sustaining ecosystems: a conceptual framework*. San Francisco, Calif., USDA Forest Service, Pacific Southwest Region.

MARSTON, R.A.; GIREL, J.; PAUTOU, G.; PIÉGAY, H.; BRAVARD, J.P. and ARNESON, C. 1995. Channel metamorphosis, floodplain disturbance, and vegetation development in a river in France. *Geomorphology*, Vol. 13, pp. 121–31.

MARTIN, J.L. and MCCUTCHEON, S.C. 1999. *Hydrodynamics and transport for water quality modelling*. Boca Raton, Fla., Lewis.

MOSSELMAN, E. 1989. Theoretical investigation on discharge-induced river-bank erosion. *Communications on Hydraulic and Geotechnical Engineering*. Delft, the Netherlands, Delft University of Technology. No. 89.3.

MOSSELMAN, E. 1992. Mathematical modelling of morphological processes in rivers with erodible cohesive banks. *Communications on Hydr. and Geotech. Engrg*. Delft University of Technology. No. 92.3.

MOSSELMAN, E. 1995. A review of mathematical models of river planform changes. *Earth Surface Processes and Landforms, BGRG*, Vol. 20, No. 7, pp. 661–70.

MOSSELMAN, E.; HUISINK, M.; KOOMEN, E. and SEYMONSBERGEN, A.C. 1995. Morphological changes in a large braided sand-bed river. In: E.J. Hickin (ed.), *River geomorphology*. Chichester, UK, Wiley, pp. 235–47.

MOSSELMAN, E. and WENSINK, G.J. 1993. River bathymetry observation with radar remote sensing. *Proceedings of the International Workshop on Morphological Behaviour of the Major Rivers in Bangladesh*, Dhaka,.November.

Dhaka, Government of Bangladesh and Commission of the European Communities, pp. 243–50.

MURRAY, A.B. and PAOLA, C. 1994. A cellular model of braided rivers. *Nature*, No. 371, pp. 54–7.

MURRAY, A.B. and PAOLA, C. 1997. Properties of a cellular braided-stream model. *Earth Surface Processes and Landforms, BGRG*, Vol. 22, pp. 1001–25.

NAIMAN, R.J.; LONZARICH, D.G. BEECHE, T.J. and RALPH, S.C. 1992. General principles of classification and the assessment of conservation potential in rivers. In: P.J. Boon, P. Calow and G.E. Petts (eds.). *River conservation and management*. Chichester, UK, Wiley, pp. 93–123.

NAKAGAWA, H.; TSUJIMOTO, T. and SHIMIZU, Y. 1992. Sediment transport in vegetated bed channel. In: *Larsen and N. Eisenhauer (eds.), Proceedings of the Fifth International Symposium on River Sedimentation*, Karlsruhe, 6–10 April.

NANSON, G.C. and CROKE, J.C. 1992. A genetic classification of floodplains. *Geomorphology*, Vol. 4. pp. 459–86.

NANSON, G.C. and KNIGHTON, A.D. 1996. Anabranching rivers: their cause, character and classification. *Earth Surface Processes and Landforms*, Vol. 21, pp. 217–39.

NATIONAL ACADEMY OF SCIENCES. 1995. *Wetlands: characteristics and boundaries*. Washington, D.C., National Academy Press.

NEILL, C.R. and GALAY, V.J. 1967. Systematic evaluation of river regime. *Journal of Waterways and Harbors Division,* ASCE, Vol. 93, No. WW1, pp. 25–53.

NELSON, J.M. and SMITH, J.D. 1989. Evolution and stability of erodible channel beds. In: S. Ikeda and G. Parker (eds.), *River meandering*. Washington, D.C., AGU (Water Resources Monograph 12). pp. 321–77.

NEWSON, M.D. and NEWSON, C.L. 2000 Geomorphology, ecology and river channel habitat: mesoscale approaches to basin-wide challenges. *Progress in Physical Geography*, Vol. 24, No. 2, pp. 195–217.

NIENHUIS, P.H. 1992. Eutrophication, water management and the functioning of Dutch estuaries and coastal lagoons. *Estuaries*, Vol. 15, pp. 538–48.

NRC (National Research Council). 1983. *An evaluation of flood-level prediction using alluvial river models*. Washington, D.C., National Academy Press.

NRC (National Research Council). 1992. *Restoration of aquatic ecosystems: science, technology, and public policy*. Washington, D.C., National Academy Press.

ODGAARD, A.J. 1981. Transverse bed slope in alluvial channel bends. *Journal of the Hydraulics Division,* ASCE, Vol. 107, No. HY12, pp. 1677–94.

ODUM, E.P. 1989. *Ecology and our endangered life-support systems*. Sunderland, Mass., Sinauer.

OLESEN, K.W. 1987. *Bed topography in shallow river bends*. Communications on Hydraulic and Geotechnical Engineering. Delft, the Netherlands, Delft University of Technology. No. 87–1.

O'NEILL, R.V.; JOHNSON, A.R. and KING, A.W. 1989. A hierarchical framework for the analysis of scale. *Landscape Ecology*, Vol. 3, Nos. 3–4, pp. 193–205.

PARK, J.K. and JAMES, A. 1988. Time-varying turbulent mixing in a stratified estuary and the application to a Lagrangian 2-D model. *Estuarine, Coastal and Shelf Science*, Vol. 27, pp. 503–20.

PARK, J.K. and JAMES, A. 1989. A unified method of estimating longitudinal dispersion in estuaries. *Water Science and Technology*, Vol. 21, pp. 981–93.

PETERS, J.J. 1988. Études récentes de la navigabilité. In: *Symp. L'accès maritime au Zaïre*, Bruxelles, 5 Décembre 1986. Académie Royale des Sciences d'Outre-Mer, pp. 89–110.

PETERSEN, M.S. 1986. *River engineering*. Englewood Cliffs, N.J., Prentice-Hall.

PHILLIPS, J.D. 1995. Biogeomorphology and landscape evolution: the problem of scale. *Geomorphology*, Vol. 13, pp. 337–47.

PIÉGAY, H. 1997. Interactions between floodplain forests and overbank flows: data from three piedmont rivers of southeastern France. *Global Ecology and Biogeography Letters*, Vol. 6, Floodplain forests special issue, pp. 187–96.

PILKEY, O.H. and DIXON, K.L. 1996. *The Corps and the shore*. Washington, D.C., Island Press.

POINCARÉ, J.H. 1914. *Science et méthode*. Paris, Flamarion, 1908. (First Book, Chapter 4, Section II.) Authorized German Edn., *Wissenschaft und Methode*. Teubner, F.L. Lindeman.

PONCE, V.M. 1989. *Engineering hydrology: principles and practices*. Englewood Cliffs, N.J., Prentice-Hall.

PRITCHARD, D.W. 1955. Estuarine circulation patterns. *Proceedings of the American Society of Civil Engineers*, No. 81, pp. 1–11.

PRITCHARD, D.W. 1967. What is an estuary? Physical viewpoint. In: G.H. Lauff (ed.), *Estuaries*. Washington, D.C., American Association for the Advancement of Science (Publ. No. 83), pp. 3–5.

PRITCHARD, D.W. 1989. Estuarine classification: a help or a hindrance. In: B.J. Neilson, A. Kuo and J. Brubaker (eds.), *Estuarine Circulation*. Clifton, N.J., Humana. pp. 1–38.

PROSSER, I.P.; DIETRICH, W.E. and STEVENSON, J. 1995. Flow resistance and sediment transport by concentrated overland flow in a grassland valley. *Geomorphology*, Vol. 13, pp. 71–86.

RADEMAKERS, J.M.G. and WOLFERT, H.P. 1994. *Het Rivier-Ecotopen-Stelsel; een indeling van ecologisch relevante ruimtelijke eenheden ten behoeve van ontwerp- en beleidsstudies in het buitendijkse rivierengebied*. Publikaties en rapporten van het project 'Ecologisch Herstel Rijn en Maas', No. 61. Lelystad, RIZA.

RICHARDS, K.S. 1982. *Rivers: form and process in alluvial channels*. London, Methuen.

RICHARDSON, E.V. 1997. Sedimentation. In: A.K. Biswas (ed.), *Water resources: environmental planning, management and development*. New York, McGraw-Hill, pp. 73–98.

RICKLEFS, R.E. 1990. *Ecology*. New York, W.H. Freeman.

RILEY, A.L. 1998. *Restoring stream in cities: a guide for planners, policy-makers, and citizens*. Washington, D.C., Island Press.

ROBERSON, J.A. and CROWE, C.T. 1996. *Engineering fluid mechanics*, 6th edn., New York, Wiley.

RUST, B.R. 1978. A classification of alluvial channel systems. In: A.D. Miall (ed.), *Fluvial Sedimentology*. Calgary, Canadian Society of Petroleum Geologists (Memoir 5), pp. 187–98.

SARKER, M.H. 1996. *The morphological processes in the Jamuna River*. M.Sc. thesis H.H.290. Delft, IHE.

SCHUMM, S.A.; HARVEY, M.D. and WATSON, C.C. 1984. *Incised channels: morphology, dynamics and control*. Littleton, Colo., Water Resources Publications.

SEEHORN, M.E. 1992. Stream habitat improvement handbook. Atlanta, Ga., US Department of Agriculture, Forest Service (Technical Publication R8-TP 16).

SHIMIZU, Y. and ITAKURA, T. 1985. *Practical computation of two-dimensional flow and bed deformation in alluvial streams*. Sapporo, Japan, Hokkaido Development Bureau (Civil Engineering Research Institute Report).

SHIMIZU, Y. and ITAKURA, T. 1989. Calculation of bed variation in alluvial channels. *Journal of Hydraulic Engineering*, ASCE, Vol. 115, No. 3, pp. 367–84.

SILVESTER, R. and HSU, J.R.C. 1997. *Coastal stabilization, advanced series on ocean engineering*, Vol. 14, Singapore, World Scientific Publications.

SIMONS, D.B. and SENTÜRK, F. 1992. *Sediment transport technology*. Littleton, Colo., Water Resources Publications.

STANLEY, S.J. and SMITH, D.W. 1992. Lagoons and ponds. *Water Environment Research*, Vol. 64, No. 4, June, pp. 367–71.

STARKEL, L. 1990. Fluvial environment as an expression of geoecological changes. *Zeitschrift fur Geomorphologie*, Suppl.-Bd. 79. Gebruder Borntraeger, pp. 133–52.

STATZNER, B. and HIGLER, B. 1985. Questions and comments on the river continuum concept. *Canadian Journal of Fisheries Aquatic Sciences*, Vol. 42, pp. 1038–44.

STRUIKSMA, N.; OLESEN, K.W.; FLOKSTRA, C. and DE VRIEND, H.J. 1985. Bed deformation in curved alluvial channels. *Journal of Hydraulic Research*, IAHR, Vol. 23, No. 1, pp. 57–79.

THOMANN, R.V. and MUELLER, J.A. 1987. *Principles of surface water quality modelling and control*. New York, Harper and Row.

THORNE, C.R. 1990. Effects of vegetation on riverbank erosion and stability. In: J.B. Thornes (ed.), *Vegetation and erosion: processes and environments* Chichester, UK, Wiley, pp. 125–44.

THORNE, C.R.; HEY, R.D. and NEWSON, M.D. (eds.). 1997. *Applied fluvial geomorphology for river engineering and management*. Chichester, UK, Wiley.

THORNE, C.R.; RUSSELL, A.P.G. and ALAM, M.K. 1993. Planform pattern and channel evolution of the Brahmaputra River, Bangladesh. In: J.L. Best and C.S. Bristow (eds.), *Braided rivers*. London, Geological Society (Spec. Publ. No. 75), pp. 257–76.

THORNES, J.B. 1990 The interaction of erosional and vegetational dynamics in land degradation: spatial outcomes. In: J.B. Thornes (ed.). *Vegetation and erosion: processes and environments*. Chichester, UK, Wiley, pp. 41–53.

TICKNER, D.; ARMITAGE, P.D.; BICKERTON, M.A. and HALL, K.A. 2000. Assessing stream quality using information on mesohabitat distribution and character. *Aquatic Conservation: Marine and Freshwater Ecosystems*, No. 10, pp. 179–96.

TOOTH, S. and NANSON, G.C. 1999. Anabranching rivers on the Northern Plains of arid central Australia. *Geomorphology*, Vol. 29, pp. 211–33.

TORREY, V.H. III. 1995. Retrogressive failures in sand deposits of the Mississippi River. In: C.R. Thorne, S.R. Abt, F.B.J. Barends, S.T. Maynord and K.W. Pilarczyk (eds.). *River, coastal and shoreline protection; erosion control using riprap and armourstone*. Chichester, UK, Wiley, 1995, pp. 361–77.

TOWNSEND, C.R. 1996. Concepts in river ecology: pattern and process in the catchment hierarchy. title *Archiv fur Hydrobiologie, Supplement 113, Large Rivers*, No. 10, pp. 3–21.

TRIMBLE, S.W. and MENDEL, A.C. 1995. The cow as a geomorphic agent; a critical review. *Geomorphology*, Vol. 13, pp. 233–53.

TSUJIMOTO, T. 1999. Fluvial processes in streams with vegetation. *Journal of Hydraulic Research*, Vol. 37, No. 6, pp. 789–803.

TURNER, M.G.; DALE, V.H. and GARDNER, R.H. 1989. Predicting across scales: theory development and testing. *Landscape Ecology*, Vol. 3, Nos. 3–4, pp. 245–52.

TURNER, R.K. and BATEMAN, I.J. (eds.). 2001. *Water resources and coastal management*. Cheltenham, UK, Edward Elgar.

USEPA (United States Environmental Protection Agency). 1995. *Ecological restoration: a tool to manage stream quality*. Washington, D.C., USEPA.

USEPA (United States Environmental Protection Agency). 2002. Wetlands, oceans and watersheds. www.epa.gov/owow (accessed 13 November 2004).

VAN BENDEGOM, L. 1947. Some considerations on river morphology and river improvement. *De Ingenieur*, Vol. 59, No. 4, (in Dutch; English transl.: National Research Council of Canada, Tech. Translation 1054 1963).

VANNOTE, R.L.; MINSHALL, G.W.; CUMMINS, K.W.; SEDELL, J.R. and CUSHING, C.E. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, Vol. 37, No. 1, pp. 130–7.

VANONI, V.A. (ed.). 1975. *Sedimentation engineering*. New York, American Society of Civil Engineering Press.

WARD, J.V.; TOCKNER, K. and SCHIEMER, F. 1999. Biodiversity of floodplain river ecosystems: ecotones and connectivity. *Regulated Rivers: Research and Management*, Vol. 15, pp. 125–39.

WILLIAMS, G.P. and WOLMAN, M.G. 1984. *Downstream effects of dams on alluvial rivers*. Washington, D.C., US Geological Survey (Professional Paper 1286).

WOLMAN, M.G. 1954. A method of sampling coarse river bed material. *Transactions of the American Geophysics Union*, Vol. 35, No. 6, pp. 951–6.

WOLMAN, M.G. and LEOPOLD, L.B. 1957. *River flood plains: some observations on their formation*. Washington, D.C., USGS (Professional Paper 282C).

YANG, C.T. 1996. *Sediment transport theory and practice*. New York, McGraw-Hill.

ZEDLER, J. 1996. Ecological issues in wetland mitigation: an introduction to the forum. *Ecological Applications*, Vol. 6, No. 1, pp. 33–7.

YEH, G.T.; CHENG, H.P.; CHENG, J.R. and LIN, H.C.J. 1998. *A numerical model simulating flow, contaminant, and sediment transport in watershed systems*. Vicksburg, Miss., USACE (Waterways Experiment Station Technical Report CHL-98-15, WASH12D).

ZWOLINSKI, Z. 1992a. Sedimentology and geomorphology of overbank flows on meandering river floodplains. *Geomorphology*, Vol. 4, No. 6, pp. 367–79.

ZWOLINSKI, Z. 1992b. Sedimentology and geomorphology of overbank flows on meandering river floodplains. In: G.R. Brakenridge and J. Hagedorn (eds.), *Floodplain Evolution. Geomorphology*, Special Issue, Vol. 4, No. 6, pp. 367–79.

# Appendix B: Monitoring and Adaptive Management

1. Introduction 559

2. System Status 561
   2.1. System Status Indicators 562

3. Information Needs 562
   3.1. Information Objectives and Priorities 563

4. Monitoring Plans 563

5. Adaptive Monitoring 564
   5.1. Risk Assessments For Monitoring 564
   5.2. Use of Models 565

6. Network Design 565
   6.1. Site Selection 566
   6.2. Sampling/Measurement Frequencies 566
   6.3. Quality Control 566
   6.4. Water Quantity Monitoring 567
   6.5. Water Quality Monitoring 568
   6.6. Ecological Monitoring 569
   6.7. Early-Warning Stations 569
   6.8. Effluent Monitoring 570

7. Data Sampling, Collection and Storage 570
   7.1. Overview 570
   7.2. Remote Sensing 571
      7.2.1. Optical Remote Sensing For Water Quality 571
      7.2.2. Applications in the North Sea 572

8. Data Analyses 572

9. Reporting Results 573
   9.1. Trend Plots 573
   9.2. Comparison Plots 573
   9.3. Map Plots 576

10. Information Use: Adaptive Management 576

11. Summary 578

12. References 578

# Appendix B: Monitoring and Adaptive Management

Monitoring the impacts or outcomes of any water management policy provides a way of assessing just how well the policy meets expectations. Developing a monitoring plan requires the identification of performance indicators and how frequently and accurately they will be measured over time and space. The major challenge in monitoring is to make the information obtained fit the information needed and then to act on it, as and when appropriate. This is called adaptive management: what most of us do throughout our lives. Over time managers should be able to improve their management and their monitoring policies on the basis of what they learn about the system they are managing. Just how much and how well they learn will largely depend on the effort given to developing and implementing an effective monitoring and adaptive management strategy.

## 1. Introduction

Monitoring is the process of observing what is happening. Managers of water resources systems need to know what is taking place in their systems, both over time and over space. This usually requires sampling one or more elements or features of the system according to pre-arranged schedules, using comparable methods for data measuring or sensing, recording, collection and analysis. Monitoring provides information on the state of the system. Adaptive management is the action taken in response to that information with the aim of improving how the system performs.

Adaptive management can be active or passive. Active adaptive management, defined by Holling (1978), involves managing the system in ways that maximize the understanding obtained from the monitored data. Active adaptive management involves experiments aimed at discovering the limits of system vulnerability and resilience. Such management can be risky. It is possible for management experiments to degrade the system, for example by killing some of the species in an ecosystem that is managed to protect them. For this reason managers are often reluctant to assume those risks just to learn more. Hence, if adaptive management is implemented, it is often passive adaptive management, sometimes called adaptive assessment or adaptive implementation. Passive adaptive management strategies strive to manage the system in the best way possible, correcting mistakes and implementing changes on the basis of what is learned as it is learned, but without making possibly damaging experiments.

The questions to be addressed in establishing a monitoring programme for adaptive management are what to measure, where and how often to measure or record each parameter, with what accuracy, and why. The challenge is to know when the information derived from a monitoring programme is sufficient to act upon, as well as what action to take. Often the results, especially the ecological impacts, of a change in a management policy will be observable only long after the change has been implemented.

These questions are addressed in the various stages of a monitoring programme's life cycle, as shown in Figure B.1. The process of monitoring and adaptive management is more than just measuring system

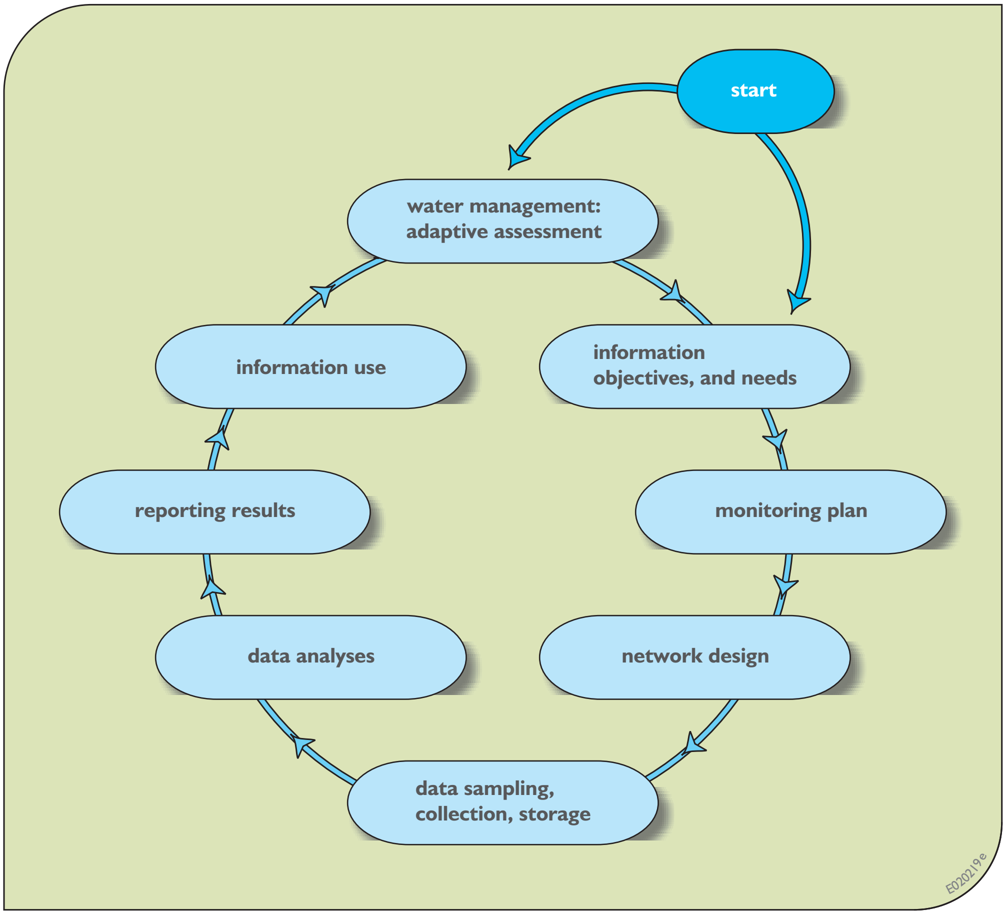



**Figure B.1.** Chain of activities in monitoring and assessment (Adriaanse and Lindgaard-Jorgensen, 1995).

attributes, collecting, storing, analysing and publishing the measured data, and then acting on the results. It is a sequence of related activities that starts with the identification of information needs, and ends with the use of the information. All too often time and money is invested to obtain data before sufficient thought has been given to how those data will or could be used and their value compared to their cost.

Modelling can help identify just what data are needed for the decisions being considered, and how accurate those data need be. (Some modelling examples in Chapters 4 and 7 illustrate this.) A dilemma, of course, is that if data obtained from a current monitoring programme are intended to be of value to future managers, it is difficult to know today just what data and what precision those future managers may want.

The design of a monitoring system starts with defining the information needed for decision-making. The information needed determines the attributes to be measured – the types of data to be collected and the kinds of analyses to be applied to them. The monitoring plan specifies these data, their required accuracy and their frequency of measurement. Frequency of measurement and the density of monitoring sites are in part dependent on the variability of an attribute's or parameter's value over time and/or space, and just how important it is to capture this temporal or spatial variability.

Once the network design has been defined, data collection, storage and analysis procedures need to be specified, along with plans for reporting and disseminating the results. This information should be included in the monitoring plan.

The last element in Figure B.1, 'information use', is input to the 'managers' of the system. Actions taken to manage the system more effectively on the basis of this new information may lead to changes in the information needs. As information needs change, this chain of activities will repeat itself. Each component of the monitoring cycle is subject to change and enhancement over time, reflecting changes in knowledge or goals, improvements in methods and instrumentation, and budgets.

## 2. System Status

Information needs depend on the issues and problems facing water managers. To identify management priorities several activities are needed. As suggested in Figure B.2, these include identifying the functions and the uses served by the system being managed. This in turn involves carrying out inventories and assessments of available and accessible information, making field surveys if information is lacking, identifying criteria and targets, and evaluating the use, costs and benefits of additional data.

Specifications of information needs should be based on the analysis of water management issues and opportunities. These in turn are determined from inventories, surveys, stakeholder concerns about what needs attention, and failures to meet standards, targets, or management criteria. Issues and targets can include existing or future problems or threats, e.g. flooding, toxic contamination, water supply shortage. They can include the full range of qualitative and quantitative aspects in multipurpose system management (see Table B.1).

Additional surveys and monitoring are needed if sufficient data are not available to identify the causes of known problems. Surveys and monitoring generate new data that can relate to a broad range of subjects, such as the evaluation of site conditions (e.g. post-flood surveys), the variability of monitoring parameters in space and time, or the screening of the occurrence of pollutants or toxic effects in water and sediments.

| functions & uses \ issues | flooding | scarcity | erosion/sedimentation | biodiversity | continuity | salinity | acidification | pollution | eutrophication |
|---|---|---|---|---|---|---|---|---|---|
| human health | ● | ● | ● | | | ● | | ● | ● |
| ecosystem functioning | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| fisheries | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| recreation | ● | ● | ● | ● | ● | ● | | ● | ● |
| drinking water | ● | ● | ● | | | ● | ● | ● | ● |
| irrigation | | ● | | | | ● | | ● | ● |
| industrial use | | ● | ● | | | ● | | ● | ● |
| hydro power | | ● | ● | | ● | | | | |
| transport medium & navigation | ● | ● | ● | | ● | | | | |

E020219g

**Table B.1.** System functions and uses.

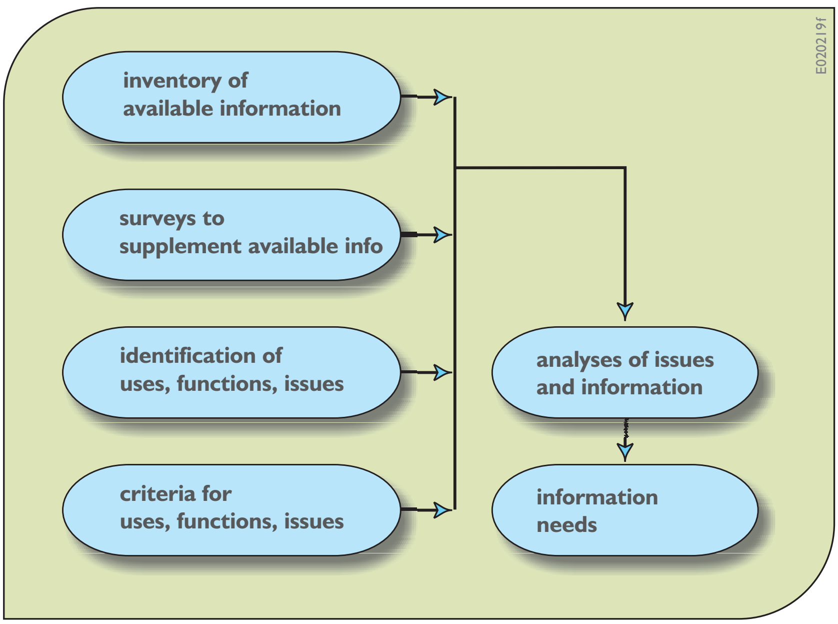



**Figure B.2.** Water management activities to identify information needs.

Water quality surveys can give additional insight into the functioning of the aquatic ecosystem and the incidence of pollution and toxic effects in the water. Investigating the structure of the macro-invertebrate community, the upstream and downstream differences in the river reach, and the changes that occur over time can provide an assessment of the biological quality of the aquatic zone of a river. Chemical screening of surface water, sediment and effluents at hot spots and key locations can allow an assessment of the chemical quality of the aquatic zone of a river. Additionally, specific target compounds, toxic effects in surface water, sediments and effluents that might be expected can be measured and analysed. By means of ecotoxicological tests, the concentrations of a broad range of chemicals and variation in sensitivity among species can be assessed.

### 2.1. System Status Indicators

System status or performance indicators should ideally be measurable parameters. Indicator parameters should sufficiently characterize or represent functions and uses of water bodies and/or be of value for testing the effectiveness of management decisions. For example, total phosphorus can be a good indicator of the eutrophication status of the river. The presence of salmon is an indicator for the ecological state of rivers such as the Columbia and Rhine Rivers. Such indicators should include those that the public care about and thus suitable for communicating with policy makers or the public. Important criteria for choosing an indicator are:

- *Communication*. The indicator should be appealing to those who will use it.
- *Simplification*. The indicator should provide insight into the situation, without having to go into great detail. The oxygen concentration in a river, for example, is a good indication of its quality for aquatic organisms; there is no need to identify the concentrations of all oxygen-demanding substances.
- *Data availability*. Sufficient data for the indicator should be available, otherwise its information content may be unreliable.

## 3. Information Needs

As many monitoring programmes are 'data rich, but information poor,' attention should be directed towards the end product of monitoring: information. The ultimate goal of monitoring is to provide information needed to answer specific management questions. Thus, a critical step in developing a successful, tailor-made and cost-effective monitoring programme is the clear definition and specification of information needs. These must be known before design criteria can be derived for the development or planning of a monitoring and assessment system.

Information needs come from management objectives. Management objectives might be stated in terms like '20% reduction of pollution in the next five years', or 'No more interruptions in the intake of drinking water within two years'. There should be an element of relativity ('percentage reduction of …') or quantity ('no more …' or 'less than …') in the specification. Next to that, an element of time ('… within two years') is imperative. Consider for example the 'Salmon 2000' slogan that summarized the efforts to restore the ecology of the River Rhine this past half decade. 'Getting the salmon back into the Rhine in the year 2000' is a goal that can be measured. There is an element of quantity (a more or less stable population of salmon) and an element of time (the year 2000). These elements can be converted into a monitoring strategy. The effectiveness of a monitoring network can only be tested when the information needs related to one or more management objectives are defined.

Five different approaches to defining information needs can be distinguished:

- *The effect approach*. There may be an adverse effect of some kind that should be reduced within a certain period. The element of relativity can be used here.
- *The source approach*. There are sources that cause adverse impacts. Their effects have to be reduced, for example by reducing their loads on the environment. This is closely related to the effect approach.
- *The achievement approach*. There is a goal to be achieved within a given time period. This approach gives an impression of the effects of the intended actions after a stated time. 'Salmon back in the River Rhine In the year 2000' is an example of this approach.

- *The background approach*. 'There may be no change in' a given parameter, or 'the river has to be back in its original state by' a given time. This is usually comparable to the ecological function of a water resource system.
- *The function approach*. The water system has to fulfil a specific function, such as being fit for salmon and/or swimming.

The different approaches listed above are often interrelated. Nevertheless, each of them can help facilitate the identification and specification of information needs. These needs should be identified and described in sufficient detail to ensure the design of a monitoring and assessment system that will meet them. Examples of such specified information needs include:

- The identification of appropriate parameters and/or indicators.
- The definition of criteria for assessment, such as considerations for the setting of standards or criteria for the choice of alarm or trigger conditions for early warning in the event of floods or accidental pollution.
- Requirements for reporting and presentation of the information, e.g. visualization, degree of aggregation, indices.
- Relevant error margins specified for each monitored indicator. What detail is relevant for decision making?
- Response times identifying when specified information is needed. In early-warning procedures, information is needed within hours, whereas for trend detection information is needed weeks or even months after sampling.
- Reliability requirements. To what extent is false information allowed? It is often impossible or prohibitively expensive to have 100% reliability. Depending on the consequences of error, more or less reliable information should be required. Together with the relevant margins of error, these reliability requirements may be a determining factor when selecting locations, frequencies and methodologies in the design of monitoring programmes.

Information needs should be comparable between places and situations, and should be linked to specific issues, which are in turn linked to specific management needs. Interested stakeholders should be involved in the process of specifying information needs alongside institutions responsible for the management and use of water resources. Both information users and information producers should be identified and should interact closely.

### 3.1. Information Objectives and Priorities

Information objectives indicate the intended use or purpose of the information and the management concern. The information may be needed for compliance with established standards or targets, for planning, for early warning of hazards, or for scientific understanding of natural processes or impacts.

As information needs are derived from issues, the prioritization of issues leads to a prioritization of information needs. Information is mostly needed on high-priority issues. If the same information need arises from various issues, this information need should be given increasing priority. By collecting this information once, a variety of issues may be addressed.

Information objectives evolve as water management develops, targets are met or policies change. Consequently, monitoring strategies often need to be adapted to changing information needs over time. Information needs require a regular rethinking (revision) of the information strategy in order to update the concept. When revising monitoring strategies for time-series measurements, one should not neglect the need for continuity (in parameters being measured, in locations where data have been collected, in the analytical methods used and so on). This continuity is needed to detect significant and reliable trends in system performance characteristics.

## 4. Monitoring Plans

A monitoring plan provides the basis or rationale for the design of monitoring networks. Monitoring plans should specify what has to be measured (also in terms of accuracy, type 1 and type 2 errors, etc.) and why. The network design specifies how and where it should be measured. The monitoring plan should also include the data analysis and reporting procedures that in turn can influence network design requirements.

Elements of a monitoring plan are:

- the information needs that will be covered by the monitoring programme and, equally important, the information needs that will not be covered by the monitoring strategy
- the type of monitoring (physical, chemical, biological, hydrological, early warning, effluent), the indicator variables to be measured and the preconditions for selecting locations (minimum/maximum distance from border, intake point, etc.) and sampling frequencies (in terms of reliability)
- the calculation methods, and the graphical, statistical and other tools (such as indices) to be used
- the preconditions, suppositions, assumptions, and descriptions of the area, relevant industries, major demands and so on
- the organizational responsibilities for the monitoring programme
- a plan for the design and implementation of the monitoring network
- an analysis of the risks and the possible problems that can lead to the failure of the monitoring programme

Monitoring plans are the bridges between information needs and monitoring networks.

The selection of the parameters to be monitored is usually based on their indicative character, their occurrence and the hazards they present. For reasons of efficiency, the number of monitored parameters should be restricted to those whose uses are explicitly identified. The benefits derived from measuring any additional parameter should be compared to the cost incurred. Since the benefits will probably not be expressed in monetary terms, this usually has to be a qualitative comparison, using judgement.

Integrated water management involves the consideration of all aspects of a water resources system. This includes its watersheds, its aquifers, rivers, lakes, reservoirs and wetlands, its estuaries and coastal waters, its natural ecosystems, its regulatory measures for environmental media, its management and monitoring strategies, and its relations to social and economic factors. An integrated approach eschews a focus on only localized separate components of the system in isolation. Monitoring plans should reflect these interdependencies and facilitate an integrated approach to water management.

An integrated management approach includes humans as a central element in the system. This implies recognition of social, economic, technical and political factors that influence the ways in which human beings use and affect the system. These factors should be assessed because of their ultimate effect on the system's integrity. For example, in trying to restore the hydrology and ecosystem of the Everglades region of South Florida to what it was like a half century ago, one cannot ignore the addition in recent decades of some 20 million people who now make their home there for at least part of the year. The needs for reliable water supplies and flood control were much less fifty years ago, as were the pollutants discharged from municipalities, agriculture and industry. These extra people, together with their pollutants, are not going away. To manage integrated systems, managers need to know their condition and how their condition reacts to people and their activities. This in turn requires monitoring not only of water quantities and qualities and ecological indicators, but also of their major drivers: humans and their activities.

## 5. Adaptive Monitoring

One approach to monitoring when the precise level of detail or precision is not known is an adaptive stepwise or phased plan, proceeding from coarse to fine assessments. At the conclusion of each step, an evaluation can be made of whether or not the information obtained is sufficient. Such stepwise testing strategies can result in a reduction in unnecessary data collection. In general, a phased approach to monitoring, going from broad to fine and from simple to advanced, may also be cost effective. Additionally, for developing countries or countries in transition, stepwise monitoring strategies going from labour-intensive to technology-intensive methods might be appropriate. In many cases, the lack of consistent and reliable data and the lack of a baseline against which progress can be measured are additional arguments for a phased approach.

### 5.1. Risk Assessments For Monitoring

Risk assessment can help considerably in prioritizing monitoring activities. For example, consider flood protection and water quality management.

- The central question in *flood prevention* is what protection is available at what price, and what remaining risk has to be accepted by society. Risk assessment (or, more comprehensively, flood risk management that includes risk assessment, mitigation planning and the implementation of measures) will show which hydrological, meteorological and other data should be monitored or observed.
- The *quality of water* in a small, sparsely populated catchment is unlikely to pose a risk to human health. Conversely, if there are refuse dumps or industrial plants in a catchment, there may be a high risk to human health and/or aquatic ecosystems. Thus, by using risk assessment one can decide which of all the monitoring activities have higher or lower priority. Identify priorities by asking what may go wrong when insufficient information is available (because of a lack of monitoring). What loss is likely when less than optimal decisions are made because of insufficient information or money? The same questions can be asked in the design or optimization of monitoring networks. What are the consequences for decision-making if there are no or only limited results from monitoring?

Risk assessment can also be used to prioritize specific pollutants, on the basis of their physico-chemical properties and toxicity. Risk assessment, regarding both biological agents and chemical substances, can also help in setting priorities for establishing health-related monitoring and/or early-warning systems in general, and in selecting appropriate parameters for monitoring in particular. Although still to be developed, good systems will include hazard identification, dose–effect relationships, exposure assessment and risk characterization (both qualitative and quantitative).

### 5.2. Use of Models

Models (numerical, analytical or statistical) can assist in developing a monitoring and assessment plan. They can help in screening (the preliminary evaluation of) alternative policies, in optimizing monitoring network design, in assessing the effectiveness of implemented measures, and in determining the physical and health impacts on humans and ecosystems. Computer models linked with geo-referenced databases can be used to analyse the impact of proposed measures, for example by simulating the flow and water level variations in a river and on floodplains during floods. Models can play an important role in early-warning systems (flood forecasting, travel time computations in emergency warning systems in the event of accidental pollution). They can be used in addition to monitoring to help understand what, where, and how often and how accurately to monitor.

Successful mathematical modelling for planning monitoring programmes is possible only if the modelling activities are integrated with data collection, data processing and other techniques and approaches for identifying system characteristics. One cannot calibrate and verify models without data, yet even uncalibrated models can often help identify just what data are needed and how accurate they need to be for the purposes of management and decision making.

## 6. Network Design

The design and operation of monitoring networks includes the selection of attributes or parameters to be measured, the locations where they are measured, and their sampling frequencies. The network design should meet the requirements specified in the monitoring plan. The type and nature of the system being monitored should be understood (most frequently through preliminary surveys), particularly its spatial and temporal variability. Monitoring of the quality and biology of the aquatic environment should be coupled with the appropriate hydrological quantity monitoring. Finally, arrangements should be made to ensure the quality of data. This requires the periodic checking and maintenance of the monitoring network as well as the computer data management and storage system.

The design of a monitoring network is influenced by its purpose or purposes. Table B.2 summarizes some of the design considerations that will vary for different purposes.

It will often be necessary to carry out a preliminary sampling and analysis programme to obtain a better understanding of the parameters to be monitored. The objective in the survey, a better understanding of the processes, is not directly related to the information need, but will give information for the monitoring network design. For instance, a survey to find out the distance it

**Table B.2.** Network design aspects for various monitoring objectives.

| design aspects | objectives: trend detection | objectives: testing for compliance | objectives: early warning |
|---|---|---|---|
| variables | long term interest/ policy goals | included in the standard | depending on possible accident spills |
| locations | significant and independent locations | representative for the water system | hot spots for accidental spills and interests |
| reliability/ accuracy | no discontinuities in time-series | representative for the test period | reliability more important than accuracy |
| sampling frequency | relatively low, depending on dynamics of the aquatic system | relatively low, depending on dynamics of the aquatic system | high frequency, depending on dispersion of pulse inputs |
| response time | not important | less important | crucial |

E020220a

takes for the water at the confluence of two rivers to mix completely will be useful when choosing water quality sampling locations.

## 6.1. Site Selection

The desired spatial coverage of a monitoring network depends on the spatial variability of the data being measured, and how important it is to capture or measure that variability. For monitoring meteorological parameters on watersheds, the shorter the distances between adjacent sampling sites, the greater probability of measuring any spatial variation that may exist. However, more monitoring equipment will be needed and hence costs will increase. There is a tradeoff between the accuracy of the estimates of spatially varying parameter values and cost. These relationships are shown in Figure B.3.

Referring to Figure B.3, if rainfall is being recorded at the monitoring sites 1 through 9 the estimate of the average rainfall over the entire watershed would be much more accurate than if only site 2 or 8 existed, or even if both existed and were used to make that estimate. Similarly for streamflow and quality gauges a, b, and c. Whatever the number of stream gauges to be used, they should be placed where one knows significant changes are likely to occur, such as just upstream and downstream of the confluence of tributaries.

## 6.2. Sampling/Measurement Frequencies

Water quantity and quality, sediment characteristics and biota vary over time as well as space. This variation over time affects decisions about the frequency of sampling. The objectives of monitoring strongly influence the time scale of interest, e.g. long-term variations for trend detection, short-term changes for flood forecasting and early warning. The required frequencies and methods of sampling (continuous sampling, grab sampling, composite sampling, etc.) will be dictated by both the temporal variability and the monitoring objectives. The cost–accuracy tradeoff relation shown in Figure B.3 applies to temporal sampling frequency as well as to spatial coverage.

## 6.3. Quality Control

Quality control should be performed to ensure the achievement of an acceptable standard of accuracy and precision.

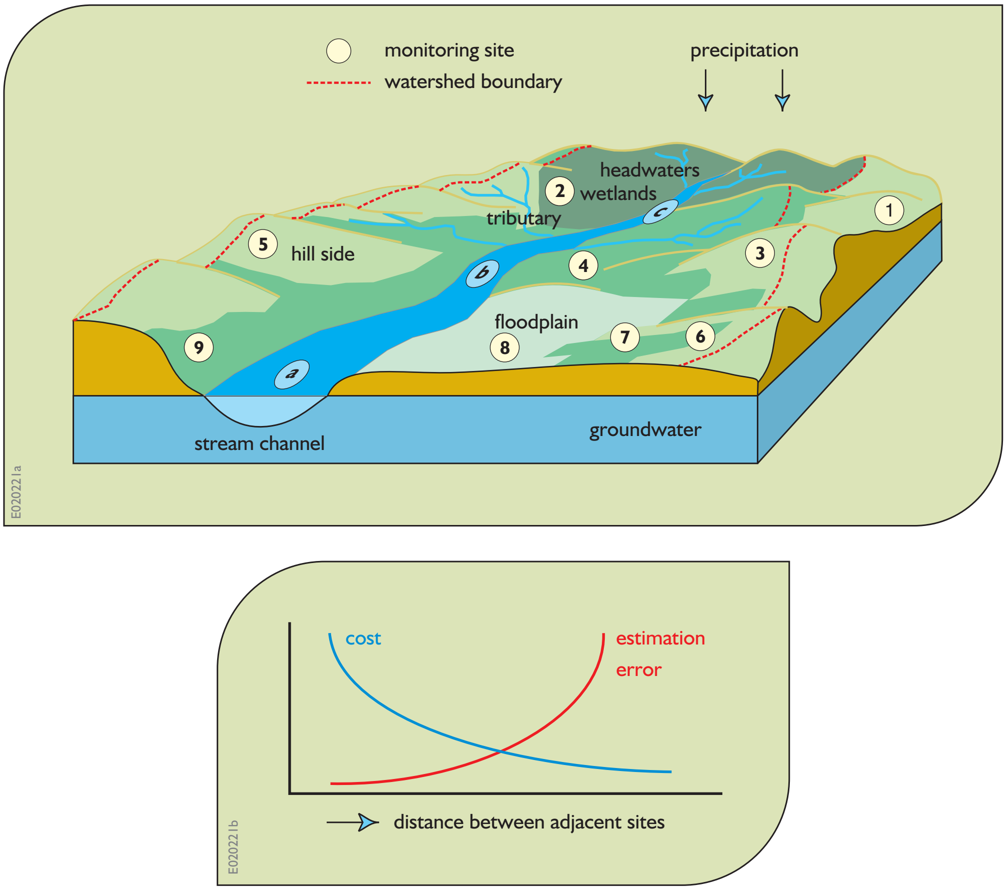



**Figure B.3. Density of monitoring sites influences estimates of average conditions over a watershed or along a stream. The density of monitoring sites also affects the cost of monitoring.**

## 6.4. Water Quantity Monitoring

The main hydrological and hydro-meteorological parameters – such as precipitation, snow cover, water level, river flow, suspended and bed load sediment discharges, evaporation and transpiration, soil moisture, and data on ice conditions – are measurable. The use of such data has increased during the past decades, due largely to developments in the use of models and forecasting systems that require these data.

Spatial representativeness is crucial in the selection of monitoring sites for hydro-meteorological parameters. The most common locations for gauging stations are the lower reaches of rivers, immediately upstream of the river mouth or where the rivers cross borders, near the confluence with tributaries and at major cities along the river. In general, a sufficient number of gauging stations should be located along the main river to permit interpolation of water level and discharge between the stations. Water balances require observation stations at small streams and tributaries as well. Gauges on lakes and reservoirs are normally located near their outlets, but sufficiently upstream to avoid the influence of drawdown.

Hydraulic conditions are an important factor in selecting sites on streams, particularly where water levels are used to compute discharge using stage-discharge rating curves. Unambiguous relationships are found at stations located on streams with natural regimes, not affected by variable backwater at the gauge caused by downstream tributaries or reservoir operations or by tidal effects.

The frequency of measurements, data transmission and forecasting depends on the variability of the hydrological characteristics and the response time requirements. Systematic water-level recordings, supplemented by more frequent readings during floods, are appropriate for most streams. The installation of water-level recorders will usually be required for streams whose levels are subject to abrupt fluctuations. For flood forecasting or flood management, telemetric systems may be used to transmit data whenever the water level changes by a predetermined amount. Continuous river flow records may be necessary in the design of water-supply systems, and in estimating the sediment or chemical loads of streams, including pollutants.

Factors to be considered in determining the number and distribution of discharge measurements within the year include:

- the stability of the stage-discharge relationship
- seasonal discharge characteristics and variability
- accessibility of the gauge in various seasons.

At new stations, many discharge measurements at different flow levels are needed to define stage-discharge relationships. At existing stations the frequency of measurements is dictated in part by the number needed to keep the stage-discharge relationship up to date. Adequate determination of discharge during floods and under ice conditions can be difficult, but is of prime importance where applicable. In situations where the channel shape can readily change during high-flow conditions, keeping an up-to-date stage-discharge relationship is a challenging (or perhaps indeed unattainable) goal.

## 6.5. Water Quality Monitoring

When water quality is monitored, water quantity (flow) should also be monitored at the same sites and times.

For specific human uses, standards may dictate the water quality parameters that need monitoring. Management issues may also dictate the parameters of interest. For ecological functioning, parameters are specified by the selected method of assessment (indices, habitat factors) and regional reference communities. The selection of hazardous pollutants as monitoring parameters will usually depend on the specific problem substances produced and/or discharged into the water and on their probability of occurrence. In practice, this should be based on results of site-specific preliminary surveys.

Nationally and internationally recognized lists of problem substances can often be used as the starting point for the selection of monitoring parameters. They are among or indicative of the pollutants that are of general concern. The availability of reliable and affordable analytical and measurement methods may also influence the selection of monitoring parameters.

In general, the selection of sampling sites is based on how representative they are. The distance between sampling locations can be critically evaluated from their degree of correlation by statistical analysis of time series of parameters. However, this is possible only if these time series are available. In transboundary rivers, sampling should preferably be performed at or near to border crossings. Sampling in the river and in the main tributaries upstream of the confluence can show the contribution (e.g. pollution load) of different tributaries. The selection of sampling sites downstream of a confluence should avoid the uncertainties related to incomplete mixing. Mixing zones can be several kilometres long, depending on the width–depth ratio and the turbulence of the main river.

Considerations of the local representativeness of the sampling point at the river site may be based on preliminary surveys, taking into account the hydrology and morphology of the river. In general, locations in the main flow of the river will be chosen for water and suspended solid sampling. Bottom sediment can best be sampled in regions where the suspended material settles. As a consequence, most sediment samples are taken near riverbanks and in the downstream sedimentation area.

The number of sampling points for sediment monitoring strongly depends on the objectives. For trend detection, a low number of sampling points or mixing samples into composite samples can sometimes yield enough information. If spatial information is to be estimated, the number of sites will increase and composite samples will normally not be used.

The selection of the sampling frequency for surface water quality parameters should be based on the variability in parameter values, and on the statistical significance and accuracy required for specific objectives. Examples of specific objectives include trend detection, load calculation and compliance testing.

Sampling frequencies for suspended solids are very similar to surface water sample frequencies. For load calculation, a higher sampling frequency is recommended during the start of flooding periods, when the main load of suspended solids is transported. The precision of the estimates obtained by sampling at regular intervals depends mainly upon the distribution of the total load over the year. The reliability of load estimates obtained with current monitoring equipment can be improved by increasing the sampling frequency.

The need to obtain information that is integrated or differentiated over time and space should determine the selection of methods for measurement and sampling of water and sediment quality and biota. There are various possible methods, including grab sampling, depth integrated sampling, time-proportional composite sampling and space composite sampling. Monitoring of the biological status implies measurement and sampling of biotic groups. Each group requires specific sampling and measurement methods; fine mesh nets collect phytoplankton and zooplankton, for example, and waterfowl can be measured through field observations. Some indicative parameters like dissolved oxygen, pH, water temperature and redox potential are best measured *in situ*, using sensor-based instruments. Such instruments require frequent calibration.

## 6.6. Ecological Monitoring

Monitoring of ecological parameters can be carried out on the level of species, communities or ecosystems. For many purposes, monitoring the habitat of communities or ecosystems is appropriate and much less demanding than doing so on a species scale. A habitat is the place where an organism or a community of organisms lives, and comprises all living and non-living factors or conditions of the surrounding environment. Habitat descriptions consider the physical environment together with the representative floral and faunal assemblages present. They are often suitable for environmental impact assessments.

Habitat and community descriptions can be based on two types of scales. The DAFOR scale (dominant, abundant, frequent, occasional, rare), adapted from the Joint Nature Conservancy Council (UK), can be used to assess habitat types. It is designed to allow consistent recording of habitat features within and between sites. This habitat detail can be used with species lists produced from site sampling. Only one feature may be recorded as dominant, while any number of features may be recorded as abundant, frequent, occasional or rare.

Species data recorded on site can be assessed using the SACFOR scale (superabundant, abundant, common, frequent, occasional, rare). Fauna and flora species can be identified and recorded where possible. When sampling is not applied, the most abundant species observed and any noteworthy or rare species at each site can be identified *in situ*, with a further list of species expected or typical of such habitats.

## 6.7. Early-Warning Stations

Measurement systems in an early-warning station are either substance- or effect-oriented. Chemical analysis screening methods can detect increases in concentrations of specific substances. Biological early-warning systems can detect deterioration in water quality through the biological effects on fish, daphnia, algae, bacteria and other species.

The pollutants that may occur in hazardous concentrations should be monitored for early warning. Automatic *in situ* sensors can measure simple indicative parameters such as dissolved oxygen, pH, or oil. If the detection of specific micro-pollutants (e.g. pesticides) is needed, advanced, but more expensive, analytical systems based on gas chromatography with mass spectrometry (GC-MS), high-performance liquid chromatography (HPLC) can be used. Toxicological effects in organisms on various trophic levels can be measured with automated biological early-warning systems.

Early-warning equipment puts high demands on operation characteristics such as speed of analysis, identification capacity and reliability of operation. Characteristics such as the precision and reproducibility of the analysis are less critical.

Early warnings should provide enough time for emergency measures to be taken. The relation between response time (the interval between the moment of sampling and the alarm) and the travel time of the pollution plume in the river, especially in high flows, from the warning station to the site where the water is used (e.g. water intake for drinking water) influences the location of an early-warning station.

The sampling sites should obviously be chosen in such a way that no pollutants are missed in the sampled water. The measurement frequency should be determined by the expected size of pollutant plumes (elapsed time for the plume to pass the station). Dispersion of the plume occurs between the discharge location and the sampling location due to the discharge characteristics of the river. Furthermore, the frequencies should provide sufficient time to take action in the event of an emergency. Additional (intensified) sampling is recommended after the first indication of accidental pollution.

### 6.8. Effluent Monitoring

The selection of effluent monitoring parameters and their priorities can be based on risk assessments. Existing national or international priority lists of chemical substances can also be helpful. In addition to specific pollutants, an emphasis should be placed on aggregate parameters and total effluent toxicity testing.

Sampling frequencies and sampling methods for effluent discharges should be based on the amount and variability of the effluent. Surveys of restricted duration using continuous or high-frequency sampling can be performed to gain the required insight into the discharge characteristics of batch and continuous effluent generation processes. The statistical significance and accuracy required for specific objectives (e.g. for compliance testing or load calculation) can be a basis for selecting the sampling frequencies and sampling methods.

## 7. Data Sampling, Collection and Storage

### 7.1. Overview

Sampling is the first stage in the actual collection of information. Methods of sampling include spot, periodic, continuous and large-volume sampling. Which method is most appropriate will depend on the variable of interest and on the characteristics of the watershed or water body. Sampling equipment should be designed to minimize the contact time between the sample and the sampler and the likelihood of sample contamination.

There are a number of sampling decisions that must be made. The first is the choice of the precise sampling site. This can be affected by the conditions at the site, the distribution of what is being measured, and the ease of access to the sampling site with the needed equipment. Second, the frequency and time of sampling must be determined. Different time-based effects can influence this decision. Natural cycles may occur, as well as production and discharge cycles of industries or other facilities just upstream. Third, there is the choice of the sampling method. Some methods, even if recommended in the network design plan, may not be usable in particular situations, for example where the water is too shallow. Fourth, decisions must be made regarding the transporting, stabilizing and storing the samples. Fifth, quality-control procedures must be implemented. All sampling methods should be periodically tested using field-based quality-control and audit procedures specifically designed to reveal the effectiveness of the entire sampling programme, including those aspects relating to the transportation, stabilization and storage of samples prior to analysis. Finally, safety requirements have to be met.

To make monitored data rapidly and conveniently available to users, they are almost always stored in computerized data files. They include the measurement data and associated meta data. The latter identify what the measured data are, when and where they were measured, what methods of measurement or laboratory analyses were used and by whom, and so on.

Often the weakest link within the monitoring programme chain is the proper storage of data. If these data are not accessible and complete with respect to the conditions and qualifiers pertaining to their collection and analysis, or are not properly validated, then the data are not likely to satisfy any information need.

Computer hardware and software used to store and manage data must be tested, maintained and upgraded regularly. The software has to insure against data loss. It must identify the correct secondary data. Furthermore, it should perform internal checks on the measured data, such as correlation analysis and application of limit pairs. Examples of software control functions are $0 < \text{pH} < 14.0$, orthophosphate-P (total-P), dissolved heavy metals (total heavy metals), and calculation of the 10% (lower) and 90% (upper) limit pairs. The software should give users a warning when data fall outside these ranges. All such calculations should be tested for accuracy before using a computer database management system.

Clear procedures should be agreed upon for the interpretation and validation of the measurement data. These will include how to deal with:

- data limitations such as missing values
- sampling frequencies that change over the period of record
- multiple observations within one sampling period
- uncertainty in the measurement procedures
- censoring the measurement signals
- small sample sizes
- outliers (values that do not conform with the general pattern of a data set)
- measurement data rounding
- data at or below the limit of detection.

## 7.2. Remote Sensing

A variety of remote sensing techniques using aerial photography or satellite images are available for monitoring some parameters. They may be used for the identification of different vegetation types, biotopes and landscape elements. Laser-altimetry provides a useful technique for monitoring forest and grassland structure and sediment bank development.

### 7.2.1. Optical Remote Sensing for Water Quality

A number of satellite-based optical sensors can be used for monitoring of water quality. The spatial resolution of these sensors renders this method of monitoring suitable for large inland water bodies or coastal waters. For remote monitoring of smaller water bodies where higher spatial resolution is needed, aircraft-based sensors can be used.

When natural sunlight, including the visible spectrum, hits the surface of a water body, some of it may be directly reflected at the surface, some may enter the water where it is absorbed and scattered by particles, and some may be transmitted through the water. The amount of absorption and scattering are the main factors influencing the reflection of light from the water, and thus the colour of the water. The substances present in the water often have unique characteristics with respect to the absorption and scattering of light. Thus the colour of water varies with the concentration of different substances in it.

Deep ocean water has a distinct blue colour because in clear water there is a lot of scattering of blue light (wavelength of blue) compared to the other wavelengths in the visible spectrum. Light absorption is relatively low. If a water body contains algae, then the light scattering by the algae cells dominates the light scattering by water. Also, there will be more light absorption at the wavelengths of blue and red light. As a result, there is a relatively high reflection of green light and the water looks green. Due to different pigments present in different algae types, algae-dominated water bodies can also take on a brown or reddish colour. Dissolved organic substances (such as humic acids) tend to give water a yellow colour, and suspended sediments rich in organic matter (e.g. dead cells) tend to give water a brownish colour.

The amount of reflection of the different colours can be used to identify which substances are present in water and, along with advanced analysis techniques, to determine the concentrations of different substances. The substances that can be quantified on the basis of optical measurements of reflected light include algae (chlorophyll), dissolved organic carbon (humic acids) and suspended sediment.

Optical remote sensing can be applied for assessment of water quality and classification of inland and coastal waters. Sediment plumes can easily be observed and can often give a good indication of the spreading of river water (typically with high sediment concentrations) in coastal seas. The extent of eutrophication can also be monitored. With several consecutive images, the changes in water quality over time can be followed. One disadvantage of remote sensing is that the frequency of available images is inflexible, and on cloudy days no images can be collected.

By using specialized software and computer models, the measured spectral reflection data can be reworked into water quality concentrations, as indicated in Figure B.4. An important aspect in this process is to correct for atmospheric influences on the measured reflection. The measured reflection values are often very low, so processes such as atmospheric scattering of light can have a considerable influence on the measurements. It is therefore important to remove the atmospheric influence by means of 'atmospheric correction' procedures.

Sensors can measure the surface temperature as well as reflected light. The method based on the reflection of natural sunlight is called 'passive' remote sensing, as opposed to 'active' remote sensing that measures the reflection of a beam sent out by the sensor itself (e.g. radar). It is possible to convert the measured spectral reflection data from an airplane or satellite sensor into

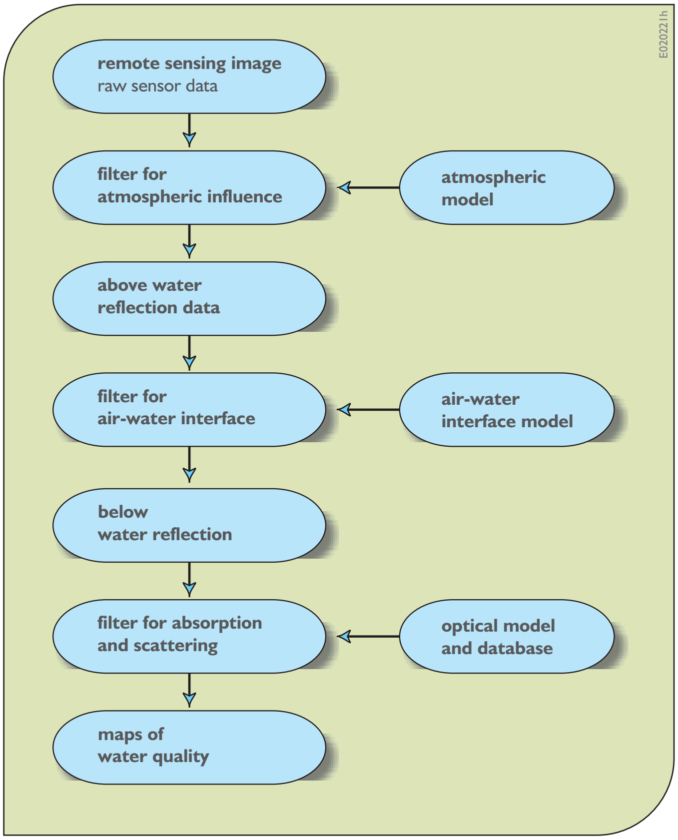



**Figure B.4.** Steps for obtaining water quality data from optical remote sensing.

an image or map of reflection values. For water quality information, the reflected light signal can be converted to concentrations of water quality parameters, such total suspended matter or chlorophyll. The steps involved in this are shown in Figure B.4.

### 7.2.2. Applications in the North Sea

The water quality of the North Sea and inland waters is based on concentrations of certain water quality parameters such as algae and suspended matter. These concentrations have traditionally been identified by collecting and analysing water samples from fixed monitoring platforms or from ships. Analyses of the samples are made in the field or at a laboratory. This method has a number of disadvantages. The spatial coverage of the sea is limited due to high costs of collecting and analysing the many samples required from each sampling site. Furthermore, the conditions at a sampling site may be changing during the time that it takes to collect the required number of samples.

Remote sensing is now being used as a relatively inexpensive monitoring method to supplement traditional water quality monitoring, by providing spatial coverage of the water body at different particular times.

## 8. Data Analyses

Data analysis converts raw data into usable information. Routine data analysis is commonly directed toward obtaining information on average conditions, trends or changing conditions, or to test for compliance with a standard. To compare and trace information obtained from the raw data, protocols for data analysis have to be developed. Using these protocols, many data analysis procedures can be automated.

Data analysis protocols should include:

- A statement of the information to be produced. This is directly related to the specified information need.
- Procedures for preparing a raw data record for graphical and statistical analysis, including how data limitations such as missing values or outliers are to be addressed before data analysis proceeds.
- Means to visually summarize the behaviour of the monitored data. Graphical presentations of the data often serve to give a better understanding of data value variability over time and space and help to interpret statistical results.
- Recommended statistical methods that yield the desired information. The selection of methods should match the statistical characteristics of the data being analysed as well as the information need.

Data analysis protocols should be established before any data are collected and analysed. Otherwise arguments can develop over the analysis methods. Statisticians can and do disagree over what statistical procedures are most appropriate. Whatever methods are used, one should understand the important assumptions that underlie them, whether these assumptions are reasonable in the particular application, and the consequences of violations of the assumptions. If a data analysis protocol is agreed upon, any subsequent discussion can focus on the resulting information.

Water quality and biological samples often require laboratory analyses. This is not the place to go into what is required for the analysis of different types of parameters, except to note that whatever analyses are performed they should be scientifically acceptable and validated. Laboratory equipment should be properly maintained and calibrated with the use of reference materials. The laboratory should undergo effective internal as well as independent quality-control audits and participate in inter-laboratory check sample schemes. Laboratory personnel should be properly trained. These and other basic elements of quality assurance should be followed and enforced to obtain reliable, verifiable results.

A major quality-control issue in data analysis is traceability. It must be possible to trace back to the raw data used in the analysis as well as to the exact analysis method. Reproduction of any previously performed analysis should lead to the same result.

Geographical information systems (GIS) are useful tools for the interpretation of spatial data such as those found on maps and satellite pictures. Integrating spatial data with time-series data, each possibly originating from different agencies/sources, into one system is not easy. Standardized interfaces should be used to interconnect databases and provide for integration with a GIS. Relational databases can be used together with GIS and data processing models. Data processing based on accepted, compatible standards will make assessment and reporting comparable, even when the software used is not the same.

# 9. Reporting Results

Selecting the method or methods of presenting the data is not a trivial issue. What methods are best depends to a large measure on the target audience. Possible presentation techniques, from a detailed presentation to an aggregated overview, include:

- Tables that list measurement data. No data are lost but the reader has to glean the needed or desired information from the data.
- Statistically-processed measurement data are transformations of the original data into values that make changes in time and/or space more visible.
- Graphs providing a medium in which, for instance, trends can be recognized at a glance. Showing standards or other references in the graph puts the system status in perspective. The amoeba-type presentation discussed in Section 9.2 is an example of this. Graphs may be line graphs, histograms, pie charts or various other forms.
- Geographically presented information shown on a map. Different data from multiple locations can be displayed as multiple layers of geographically referenced information. This often provides a better understanding of the spatial distribution of the parameters involved.
- Aggregated information for rapid interpretation of large amounts of data, for example using indices. Quality indices are often used for biological quality assessments.

## 9.1. Trend Plots

Trends are often displayed as time-series plots, examples of which are shown in Figure B.5. If a significant long-term trend exists, it may be apparent upon visual examination of a plot of the raw data. Many parameters exhibit strong seasonality and therefore a 'typical' range of values can also be shown. In Figure B.5, the 'typical' range was defined to encompass approximately 70% of the values for each month. In this case, 15% of the values should exceed the high value and 15% should be lower than low value. The shaded areas in Figure B.5 represent the 'typical' range across the middle of each plot. Comparing the data points to that shaded area makes any trend more apparent should it exist.

If seasonal variability is too great to use the time-series plots to identify long-term trends, a twelve-month moving mean can be calculated for each site. The moving mean smoothes out seasonality in many cases and makes long-term trends easier to see. If the moving mean plot suggests that a trend might exist, the raw data can then be analysed further for trends.

## 9.2. Comparison Plots

One way of comparing data is through the use of amoeba plots. An amoeba plot is a schematic representation of a given condition compared to the 'natural' average or

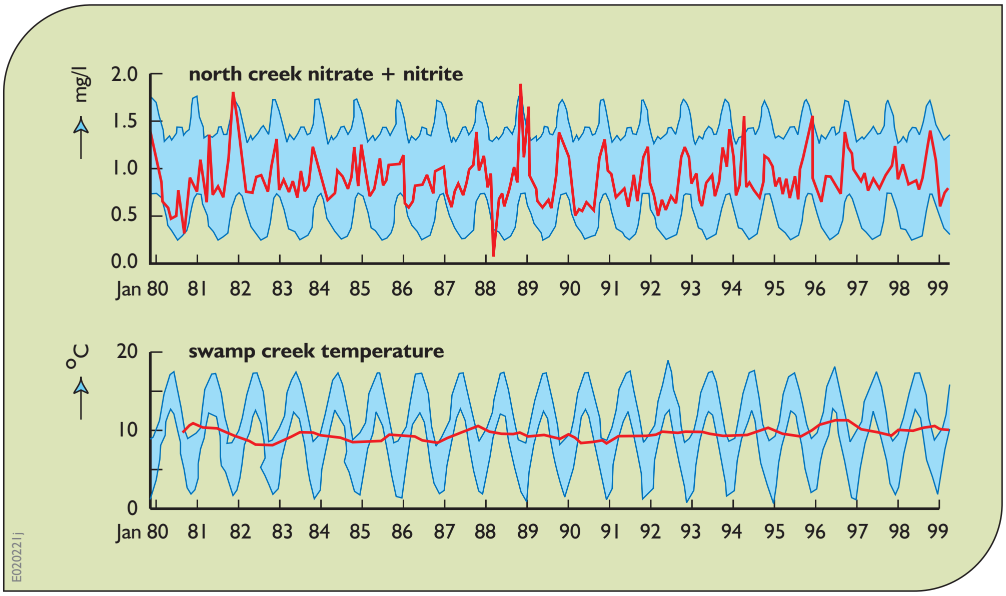



**Figure B.5.** Examples of trend plots. The shaded areas are the ranges of values in a specified percent of the data. A moving mean plot is shown in the lower graph for temperature data.

baseline condition. For the water body under study, a set of parameters considered to be representative of the water body's condition is chosen. The reference 'system' is represented by plotting the value of the parameters under 'natural' conditions on a circle. The present values of the selected parameters are plotted relative to the circle. This provides an amoeba-like figure, representing deviations from the reference or normal state, as illustrated in Figure B.6.

The figure shows stream health as indicated by eight measures of stream bugs (benthic macroinvertebrates). Stream bugs are excellent indicators of stream health. They are relatively easy and inexpensive to collect. They play a crucial role in the stream nutrient cycle, and their populations affect the whole ecosystem. The presence or absence of pollution-tolerant and intolerant bug types can indicate the condition of the stream. Population fluctuations might indicate that a change (positive or negative) has occurred in the stream. One can detect population fluctuations in a short period of time.

The circle in Figure B.6 represents the normal (healthy) value of each parameter. Deviations from that circle are expressed as percentages of these normal values. Log scales are sometimes convenient.

Amoeba plots can be used to compare data at a given site or compare the data at one site with those of other sites in a specified region.

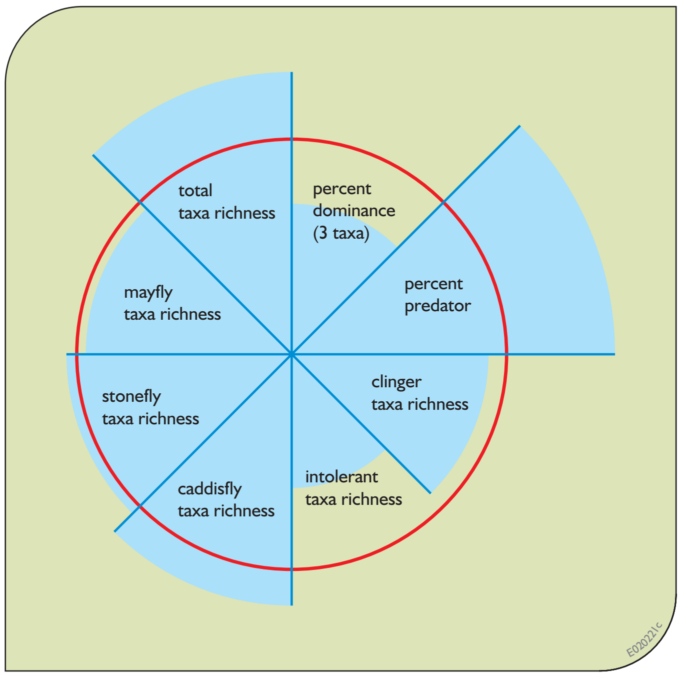



**Figure B.6.** Amoeba diagram showing status of a system with respect to the target or normal state of, in this example, eight specified parameters (bug species).

Figure B.7 illustrates another way to compare the baseline averages measured in one stream to the median levels for all tributaries measured in the region. The shaded area represents the range in which the middle 50% of all

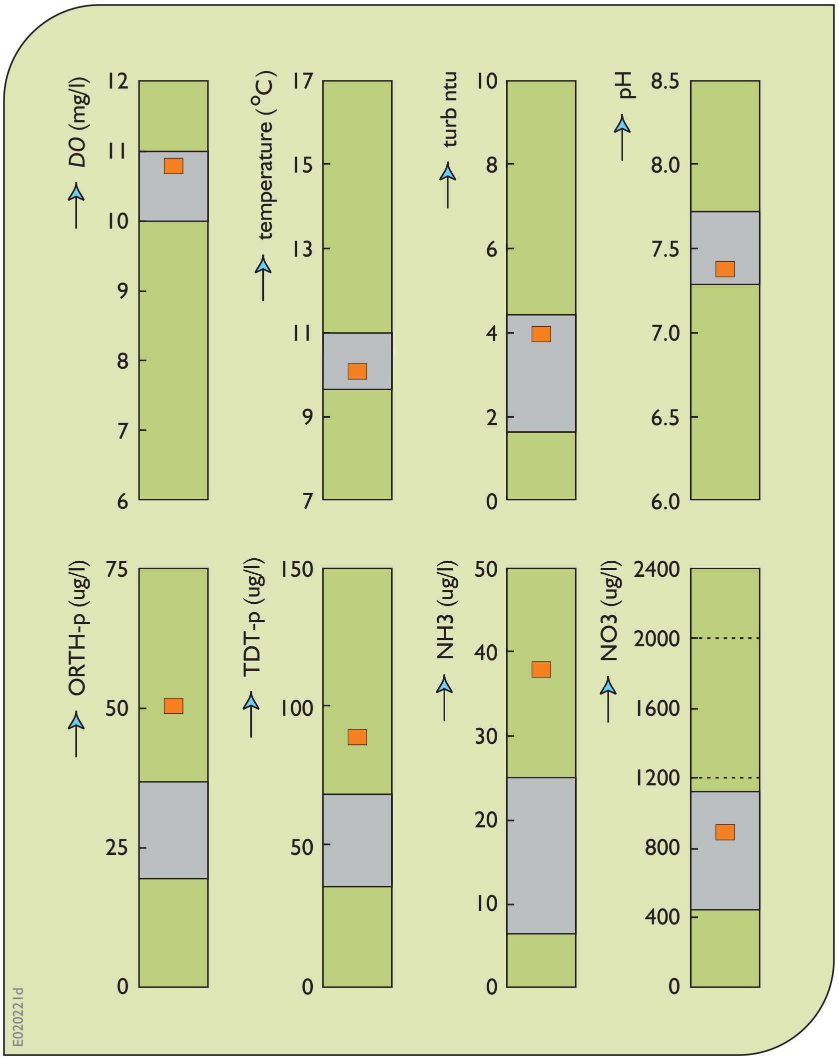



**Figure B.7.** Comparison plots showing how selected parameter values at one site compare with the values at other sites in the area. The grey bands identify the middle 50% ranges. The small red squares are the specific site values.

site averages fell, i.e. 25% were higher and 25% were lower than the shaded area. Along with these charts can be tables, not shown here, that list the average, minimum and maximum values for each parameter for each stream.

Other ways of displaying data are shown in Table B.3. In this case the data are being summarized with respect to the percentage of values that met specified standards.

Scorecards can also be used to compare data. Consider a biotic integrity index composed of ten different indicators of stream biology. Each indicator characterizes some aspect of the community that responds to degradation. The actual value of each indicator is calculated, and from that value, a score of 1, 3, or 5 is assigned to the indicator. A score of 5 indicates little or no degradation, a score of 3 moderate degradation, and a score of 1 severe degradation. The ten metric scores are then added to produce the overall score that ranges from 10 to 50. (If any indicator values are missing no score is given.) The resulting scorecard is shown in Table B.4.

Data can be interpreted by risk assessments as well. This refers to the comparison of measured, modelled or predicted values with target values. Target (desired) values of various parameters will be based on specific functions of waters, such as use for drinking water or recreation. The measured or predicted data are referred to as Predicted Values (PV), and the target or desired values are referred to as Target Values (TV). For the former, terms like Predicted Environmental Concentrations (PEC) of pollutants are sometimes used. For the latter, terms like Predicted No Effect Concentrations (PNEC) or Maximum

| parameter collected | # of samples | # of samples not meeting criteria | % of samples not meeting criteria |
|---|---|---|---|
| *D.O.* | 231 | 16 | 7.1 |
| Temperature | 245 | 17 | 6.9 |
| Turbidity | 234 | 15 | 6.4 |
| pH | 231 | 1 | 0.4 |
| Enterococcus | 127 | 64 | 50.4 |
| Fecal Coliforms | 235 | 172 | 73.2 |

E020221m

**Table B.3.** Ways of displaying information pertaining to whether or not sample data met the standards for the selected parameters.

| code | site name | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 |
|---|---|---|---|---|---|---|---|---|
| 1 | Nicki Creek | 44 | 36 | 28 | 26 | | 30 | 22 |
| 2 | Bear Creek | 36 | 34 | <30 | 28 | | 26 | 16 |
| 3 | Eelco Creek | 25 | 26 | | | | 20 | 14 |
| 4 | Simon Creek | | <30 | | dry | | dry | dry |
| 5 | Bug Creek | 35 | <30 | | 30 | | 22 | 28 |
| 6 | Jos Creek | | | | | | 28 | 24 |
| 7 | Sus Creek | | | 26 | 26 | 18 | 24 | 26 |
| 8 | Jen Creek | 32 | 22 | | | 20 | 28 | 24 |
| 9 | Beaver Creek | | <30 | | | | dry | dry |
| 10 | Wolf Creek | | <30 | | 24 | | 22 | 22 |
| 11 | Erik Creek | 18 | 26 | | | 32 | 34 | 28 |

E020221k

**Table B.4.** Scorecard for indices of biotic integrity.

Permissible Concentration (MPC) or function-related directives are also used. The risk quotient is the ratio of predicted (PV) over target (TV). This ratio will indicate the relative priority of that parameter.

The outcome of a sediment quality assessment could be expressed as a PEC/PNEC ratio. If this ratio is <1, little priority is given to the potential risk derived. If this ratio is >1, a certain risk is indicated. Classifying the responses might help to visualise the estimated risks in time trends or spatial gradients. The more function-related the quality criteria used, the more specific the conclusions that can be drawn on which function might be impeded due to the pollution present.

### 9.3. Map Plots

Two examples of map plots are shown in Figures B.8 and B.9.

E020221f

**Figure B.8.** Map display showing lake shore areas, in red, that are at risk of bank failure.

## 10. Information Use: Adaptive Management

Management decisions will always be made on the basis of uncertain information. These uncertainties in our ability to predict the impacts of our management decisions motivate the use of adaptive approaches to management. Adaptive management is the process by which management policies change in response to new knowledge gained from research and new information obtained from monitored data about the system being managed. Adaptive management requires a monitoring programme to detect changes in the system, the ability to evaluate trends in system performance and, finally, the authority and willingness to modify management decisions in response to those trends in an effort to improve system performance.

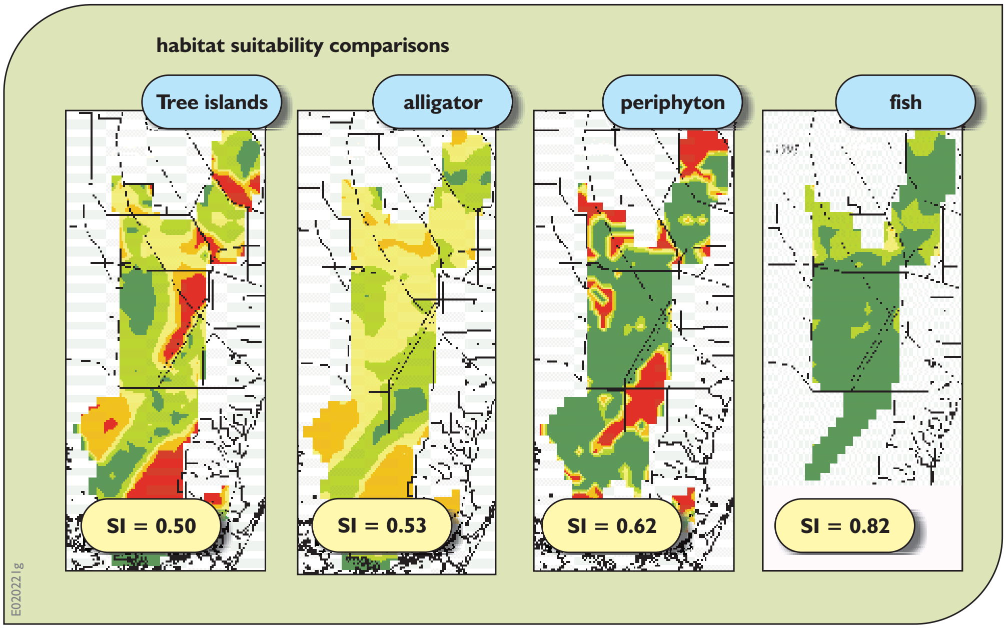



**Figure B.9.** Map displays showing distribution of habitat suitability index (SI) values for four ecosystem indicators applicable to southern Florida in the United States. Separate colors represent ranges of SI values. The average index value is shown for each indicator. (Based on Tarboton et al., 2004).

Models of the hydrological and ecological responses to management decisions, together with monitored observations, are essential components of any adaptive approach to management. An effective research strategy can lead to improved monitoring designs, improved interpretation of monitored parameter values, and improved predictive power of models and other assessment tools used in management. An integrated approach to monitoring, modelling, research and management can lead to an improved understanding of how the overall system functions and how best management practices can be implemented. Each component continually needs refining, as our understanding of the system being managed increases.

Adaptive management and decision-making is a challenging blend of scientific research, monitoring and practical management that provides opportunities to act, observe and learn, and then react. Both monitoring and management actions need to be adaptive, continually responding to an improved understanding that comes from the analysis of monitored data in comparison to model predictions and scientific research. It is a cycle, as illustrated in Figure B.10. Adaptive management requires explicit consideration of system structure and function, well-defined management goals and actions, and assessment of the anticipated system response to management decisions.

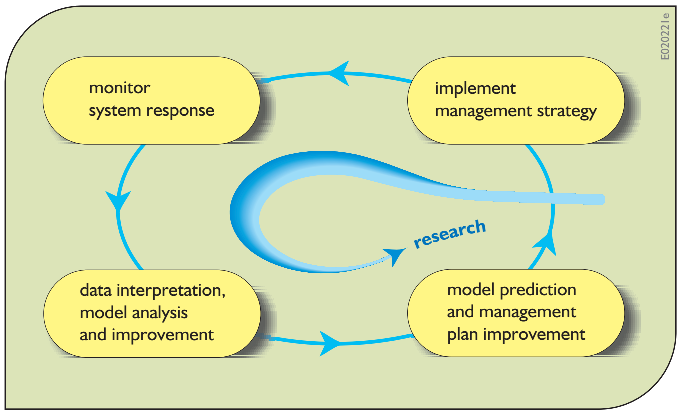



**Figure B.10.** The continually updating process of an adaptive approach to management.

The design of an adaptive management plan is best accomplished in cooperation with policy-level personnel who have the authority to commit the resources and technical personnel needed to identify scientific issues and evaluate monitoring data. Management that is to be adaptive has to rely on monitored data, and given the response times of many hydrological and ecosystem performance criteria used to judge success or failure, there must be a long-term commitment of resources to monitoring and all that it entails. But monitoring costs money and hence should be efficient. In the interest of cost savings, some water management agencies intensively monitor river basin processes and parameters only periodically, say every five years, and carry out more routine, low-intensity monitoring in the intervening period.

## 11. Summary

The degree of success of any monitoring and adaptive management programme can be expressed using the two terms 'effectiveness' and 'efficiency'. Effectiveness is the extent to which the information obtained from monitoring meets the information needs of management, and the extent to which management decisions reflect this increased information and understanding. Efficiency is concerned with obtaining the information at the least possible cost in funding and personnel. The two aspects address the two fundamental questions:

- Are the data being collecting the right data?
- Can the required information be obtained at a lower cost?

The answers to these questions are strongly related to the available budget for monitoring, and to the issue of whether more or other information is needed. If more information is required (e.g. biological monitoring in addition to chemical; or ambient water and sediment monitoring in addition to effluents) and no more money is available, then certain aspects of the existing programme must be reduced. For example, parameters reflecting the combined impacts from a number of separate parameters can be considered instead of single water quality variables. The number of stations or variables, or the frequency of sampling can be altered. If the need for other information means reductions in current data monitoring (e.g. measuring surface water quality data instead of effluent data, or using biological classification instead of chemical classification), then conflicts may arise with existing regulations and data users.

The total benefits and costs related to the monitoring information system are not only dependent on the number of sampling locations or the equipment and personnel required for laboratory analyses. Benefits also include any increased system performance due to the additional monitoring information. To quantify the benefits in financial terms, it is necessary to calculate the consequences of a shortfall in the monitoring design, for example the financial implications of incorrect decisions, or the cost related to the absence of specific information on a particular water system.

The design of a monitoring programme has to be based on the information requirements, which in turn are related to the needs of water managers. Monitoring programmes should be iterative in character. The design of the future monitoring system should be based on information collected by the existing monitoring programme. Monitoring programmes must be constantly 'designed', specified, detailed, described, or documented and updated to be sure that the system continually produces the information desired.

## 12. References

ADRIAANSE, M. and LINDGAARD-JØRGENSEN, P. 1995. *Information system for water pollution control.* Draft paper.

HOLLING, C.S. (ed.). 1978. *Adaptive Environmental Assessment and Management.* New York, Wiley.

TARBOTON, K.; LOUCKS, D.; DAVIS, S. and OBEYSEKERA, J. 2004. *Habitat suitability indices for evaluating water management alternatives.* West Palm Beach, Fla., South Florida Water Management District, (Draft Technical Report).

### Additional References (Further Reading)

ADRIAANSE, M.; NIEDERLÄNDER, H.A.G. and STORTELDER, P.B.M. 1995. *Chemical monitoring: monitoring water quality in the future*, Vol. 1. Lelystad, the Netherlands, Institute for Inland Water Management and Waste Water Treatment (RIZA).

BUISHAND, T.A. and HOOGHART, J.C. (eds.). 1986. *Design aspects of hydrological networks*. The Hague, Proceedings and Information/TNO Committee on Hydrological Research (no. 35).

CHAPMAN, D. (ed.). 1992. *Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring*. London, Chapman and Hall (published on behalf of UNESCO, WHO and UNEP).

COFINO, W.P. 1993. Quality assurance in environmental analysis. In: D. Barceló (ed.), *Environmental analysis: techniques, applications and quality assurance*. Amsterdam, Elsevier Science, pp. 359–81.

COFINO, W.P. 1994. Quality management of monitoring programmes. In: M. Adriaanse, J. van de Kraats, P.G. Stoks and R.C. Ward (eds.), *Monitoring Tailormade*. Proceedings of the international workshop, Beekbergen, the Netherlands, 20–23 September 1994. Beekbergen, RIZA, pp. 178–187.

EUROWATERNET. 1998. *Technical guidelines for implementation: The European Environment Agency's monitoring and information*. Copenhagen, Network for Inland Water Resources (Technical Report No. 7).

GILBERT, R.O. 1987. *Statistical methods for environmental pollution monitoring*. New York, Van Nostrand Reinhold.

GROOT, S. and VILLARS, M.T. 1995. *Organizational aspects of water quality monitoring: monitoring water quality in the future*, Vol. 5. Delft, the Netherlands, Delft Hydraulics.

HARMANCIOGLU, N.B.; FISTIKOGLU, O.; OZKUL, S.D.; SINGH, V.P. and ALPASLAN, M.N. 1998. *Water quality monitoring network design*. Dordrecht, the Netherlands, Kluwer Academic.

HOFSTRA, M.A. 1994. Information is vital for the national decision maker. In: M. Adriaanse, J. van de Kraats, P.G. Stoks and R.C. Ward (eds.), *Monitoring Tailormade*. Proceedings of the international workshop, Beekbergen, the Netherlands, 20–23 September 1994. RIZA, pp. 43–54.

KING COUNTY, WASHINGTON, USA. 2004. http://dnr.metrokc.gov/wlr/waterres/streams/streamsites.htm (accessed15 November).

SANDERS, T.G. and LOFTIS, J.C. 1994. Factors to consider in optimization of a monitoring network. In: M. Adriaanse, J. van de Kraats, P.G. Stoks and R.C. Ward (eds.), *Monitoring Tailormade*. Proceedings of the international workshop, Beekbergen, the Netherlands, 20–23 September 1994. RIZA, pp. 146–52.

UN/ECE. 2000. *Guidelines on monitoring and assessment of transboundary rivers*. Lelystad, the Netherlands, RIZA, Institute for Inland Water Management and Waste Water Treatment, UN/ECE Task Force on Monitoring and Assessment project-secretariat.

UN/ECE. 1993. Guidelines on the ecosystems approach in water management. In: *Protection of water resources and aquatic ecosystems*. New York and Geneva, United Nations (Water Series No. 1, ECE/ENVWA/31, United Nations Economic Commission for Europe).

WALTERS, C. 1986. *Adaptive management of renewable resources*. New York, McMillan.

WARD, R.C.; LOFTIS, J.C. and MCBRIDE, G.B. 1990. *Design of water quality monitoring systems*. New York, Van Nostrand Reinhold.

WINSEMIUS, P. 1986. *Guest in own house: considerations about environmental management*. Alphen aan de Rijn, Samson H.D. Tjeenk Willink. (In Dutch.)

## Appendix C: Drought Management

1. Introduction 581

2. Drought Impacts 581

3. Defining Droughts 584

4. Causes of Droughts 585
   - 4.1. Global Patterns 586
   - 4.2. Teleconnections 588
   - 4.3. Climate Change 588
   - 4.4. Land Use 590

5. Drought Indices 590
   - 5.1. Percent of Normal Indices 590
   - 5.2. Standardized Precipitation Index 590
   - 5.3. Palmer Drought Severity Index 591
   - 5.4. Crop Moisture Index 592
   - 5.5. Surface Water Supply Index 592
   - 5.6. Reclamation Drought Index 593
   - 5.7. Deciles 594
   - 5.8. Method of Truncation 594
   - 5.9. Water Availability Index 594
   - 5.10. Days of Supply Remaining 595

6. Drought Triggers 596

7. Virtual Drought Exercises 596

8. Conclusion 598

9. References 599

# Appendix C: Drought Management

Throughout history droughts have had far-reaching effects on humans and their civilizations. Droughts have caused crop failures and the death of natural vegetation, livestock, wildlife and people. The World Health Organization estimates that droughts and their effects cause half the deaths worldwide due to all natural disasters. Economic losses from prolonged droughts often exceed those from other more dramatic natural hazards. Humans have choices: they can increase the adverse impacts of droughts or they can take measures to reduce them. Droughts are not going to be preventable. Humans need to learn how better to live with and manage them, as with all other types of natural events.

## 1. Introduction

Droughts are normal, recurrent, yet relatively infrequent features of climate. Their timing and severity are unpredictable. They can occur virtually anywhere that precipitation occurs. Arid and semi-arid areas where precipitation is minimal or nonexistent are indeed dry, but they are not considered to be in a perpetual state of drought. Droughts occur when water supplies are 'substantially below' what is usually experienced for that place and time. Just what is considered 'substantially below' is rather arbitrary, depending on the location and what features of a drought cause the most stress or loss. For an often-cited example, a drought on the coast of Libya would result when annual rainfall is less than 180 mm. In Bali a drought might be considered to occur after a period of only a week without rain! This is why it is difficult to define droughts in a consistent way that applies to all situations. However defined, droughts are larger and longer-lasting water supply deficits than those associated with the usual short-term variations of climate.

Droughts can be related to the timing of rainfall and/or the amount or effectiveness of the rains. Other climatic factors such as high temperature, high wind and low relative humidity can significantly aggravate their severity. Figure C.1 shows the drought susceptible regions of the world and the relative levels of stress that could occur should substantially less water be available than what is considered normal.

Some areas shown as having a relative low or negligible stress in Figure C.1 can experience droughts. Figure C.2 shows the areas of the world that experienced drought conditions in 1982 and 1983.

## 2. Drought Impacts

Droughts are characterized by dry, cracked soils on river beds and lakes, dust, and thirsty plants and animals. The Australian Bureau of Meteorology provides some pictures of these impacts, shown in Figure C.3.

Damage from droughts can exceed that resulting from any other natural hazard. In the United States their annual expected impacts are estimated to exceed $6 billion. Drought primarily affects agriculture, transportation, recreation and tourism, forestry and energy sectors. Social and environmental impacts are also significant, although it is difficult to assign a monetary value to them.

When a drought begins, the agricultural sector is usually the first to be affected because of its dependence on soil moisture, which can be rapidly depleted during extended dry periods. If precipitation deficiencies continue, then users dependent on other sources of water will begin to feel

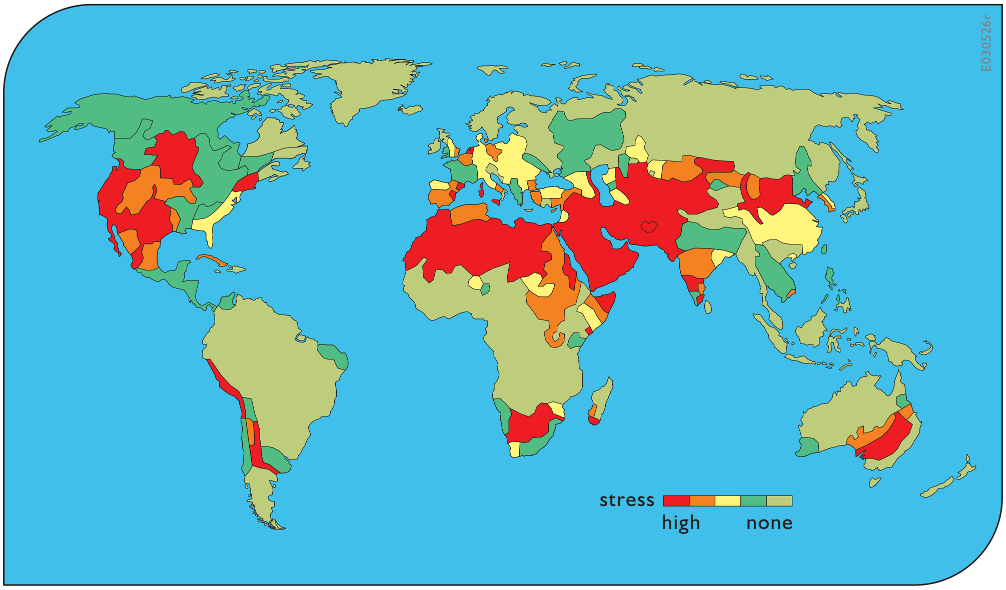



**Figure C.1.** Map showing the relative stress caused by water supply deficits should a prolonged drought occur in any of the drought-prone regions of the world.

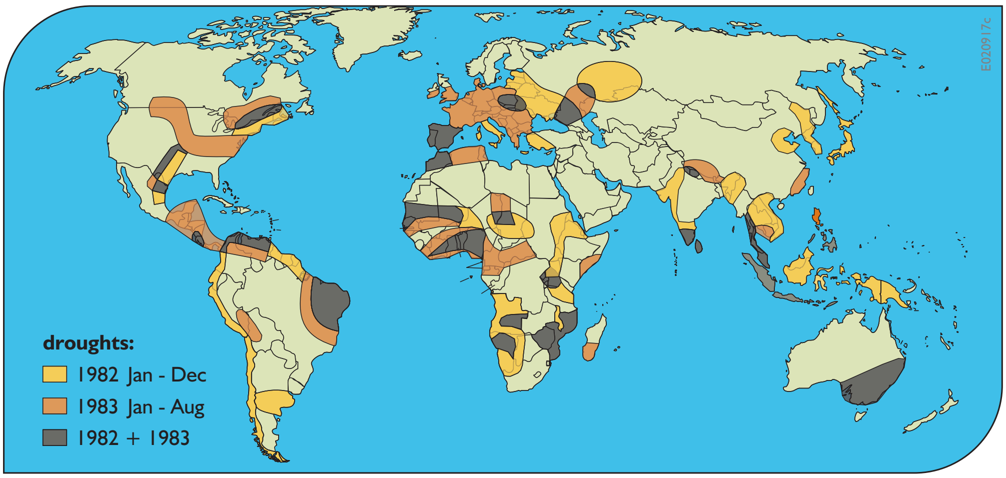



**Figure C.2.** Map showing areas of drought conditions in 1982–83 (National Drought Mitigation Center, 2000).

the effects of the shortage. Those users who rely on surface water (reservoirs and lakes) and groundwater, for example, are often the last to be affected. A short-term drought that persists for three to six months may have little impact on these water users, depending on the characteristics of the hydrological system and water-use requirements.

When precipitation returns to normal, the sequence is repeated for the recovery of surface and subsurface water supplies. Soil water reserves are replenished first, followed by streamflow, reservoirs and lakes, and groundwater. Drought impacts may diminish rapidly for agriculture because of its reliance on soil moisture, but can linger for months or even years for other users who are dependent on stored surface or subsurface supplies. Groundwater users, often the last to be affected by drought during its onset, may be the last to experience a return to normal water levels. The length of the recovery period is a function of the intensity of the drought, its duration and the quantity of precipitation received as the episode terminates.

The socio-economic impacts of a drought can follow all the other impacts just described. These impacts occur when the demand for an economic good exceeds supply. In Uruguay, for example, the 1988–89 drought resulted in significantly reduced hydropower production because power plants were dependent on streamflow rather than

**Figure C.3.** Signs of drought: dry dirt, thirsty animals and dust (courtesy of the Australian Bureau of Meteorology. http://www.k12.atmos.washington.edu/~gcg/RTN/Figures/adrought.html).

| location | major social impacts | costs |
|---|---|---|
| Mexico & Central America | n/a | $ 600 million |
| Southern Peru & Western Bolivia | n/a | $ 240 million |
| Australia | 71 dead, 8,000 homeless | $ 2.5 billion |
| Indonesia | 340 dead | $ 500 million |
| Philippines | n/a | $ 450 million |
| Southern India, Sri Lanka | n/a | $ 150 million |
| Southern Africa | disease, starvation | $ 1 billion |
| Iberian Peninsula, Northern Africa | n/a | $ 200 million |
| United States | agricultural crop loss | $ 10-12 billion |

**Table C.1.** Effects of the 1982–83 worldwide droughts (from the *New York Times*, 2 August 1983; data updated 7 February 1996).

storage for power generation. Reducing hydroelectric production required the government to use more expensive (imported) petroleum and implement stringent energy conservation measures to meet the nation's power needs.

Estimates of the economic impacts of the 1982–83 droughts, perhaps the most widespread drought event in recorded history (see Figure C.3), are given in Table C.1 (NOAA, 1994).

In North America one of the worst drought periods was from 1987–89. Economic losses from that drought in the United States exceeded $39 billion (OTA, 1993). This damage can be compared to that caused by the most costly flood, earthquake and tropical storm events.

The worst storm event in US history was Hurricane Andrew. On 24 August 1992, this 'costliest natural disaster', as it is called, hit South Florida and Louisiana. The storm killed sixty-five people and left some 200,000 others homeless. Approximately 600,000 homes and businesses were destroyed or severely damaged by the winds, waves and rain. Large parts of South Florida's communications and transportation infrastructure were significantly damaged. There was loss of power and utilities, water, sewage treatment and other essentials, in some cases up to six months after the storm ended. Andrew also damaged offshore oil facilities in the Gulf of Mexico. It toppled thirteen platforms and twenty-one satellites, bent five other platforms, and twenty-three other satellites, damaged 104 other structures and resulted in seven pollution incidents, two fires, and five drilling wells blown off location. The damage caused by Andrew in both South Florida and Louisiana totalled some $26 billion dollars.

The costliest earthquake in US history was the Loma Prieta earthquake. At 5 p.m. on 17 October 1989 the San Andreas Fault system in northern California had its first major quake since 1906. Four minutes later, as over 62,000 fans filled Candlestick Park baseball stadium for the third game of the World Series and as the San Francisco Bay Area evening commute moved into its heaviest flow, a Richter magnitude 7.1 earthquake struck. The Loma Prieta earthquake was responsible for 62 deaths, 3,757 injuries, and damage to over 18,000 homes and 2,600 businesses. About 3,000 people were left homeless. This 20-second earthquake, centred about sixty miles south of San Francisco, was felt as far away as San Diego in southern California and Reno in western Nevada. Damage and interruptions to business cost about $10 billion, with direct property damage estimated at $6.8 billion.

The most devastating flood in US history occurred in the summer of 1993. All large midwestern streams flooded, including the Mississippi, Missouri, Kansas, Illinois, Des Moines and Wisconsin rivers. The floods displaced over 70,000 people. Nearly 50,000 homes were damaged or destroyed and fifty-two people died. Over 31,000 km$^2$ (12,000 square miles) of productive farmland were rendered useless. Damage was estimated between $15 and 20 billion.

Hurricane Andrew, the Loma Prieta earthquake, and Mississippi flood events were sudden and dramatic. Droughts, on the other hand, are usually neither sudden nor dramatic; they are not given names. They can nevertheless be much more costly. As mentioned earlier, the cost of the 1988–89 drought exceeded $39 billion. Drought planning and implementing mitigation measures can help reduce those costs.

## 3. Defining Droughts

Operational definitions of droughts attempt to identify their beginning, end and degree of severity. Knowing exactly when a drought is beginning, so that water conservation measures can be implemented, is difficult since at the time it begins, whenever that is, the drought's severity and duration are unknown. Usually the beginning date is defined when some arbitrary drought index threshold is exceeded. Such a threshold might be, for example, observing less than 75% of the average precipitation in a two-year period. Alternatively, it could be based on the total volume of water in reservoirs used for water supply. The threshold or index that defines the beginning of a drought is usually established arbitrarily, depending on the hydrological characteristics of the basin and the specific demands for water in the region.

Meteorologists usually consider droughts in terms of their relative dryness and dry-period duration. What is considered abnormally dry will differ in different regions, since the atmospheric conditions that result in deficiencies of precipitation vary from region to region. Some measures of dryness identify the number of days with precipitation less than some specified threshold. This measure is appropriate for regions having a humid tropical or sub-tropical climate, or a humid mid-latitude climate characterized by year-round precipitation regimes. Examples are locations such as Manaus, Brazil; New Orleans, Louisiana (USA); and London, England. Regions such as the central United States, Europe, northeast Brazil, West Africa and northern Australia are characterized by seasonal rainfall patterns. Extended periods without rainfall are common in, for example, Omaha, Nebraska (USA); Fortaleza, Ceará (Brazil); and Darwin, Northwest Territory (Australia). In these cases a definition based on the number of days with precipitation less than some specified threshold is unrealistic. Actual precipitation departures from average amounts on monthly, seasonal or annual time scales may be more useful.

Farmers are mostly concerned about the susceptibility of crops to soil moisture deficits during different stages of crop development. Plant water demand depends on prevailing weather conditions, the biological characteristics of the specific plant, its stage of plant growth, and the physical and biological properties of the soil. Deficient topsoil moisture at planting may hinder germination, leading to low plant population densities and a reduced final yield. However, if topsoil moisture is sufficient for early growth requirements, deficiencies in subsoil moisture at this early stage may not affect final yield, provided subsoil moisture is replenished as the growing season progresses or rainfall meets plant water needs.

A drought threshold for farmers might involve daily precipitation values together with evapotranspiration rates that are used to determine the rate of soil moisture depletion and its effect on plant behaviour (growth and yield) at various stages of crop development. This measure of drought severity could be tracked over time, continually re-evaluating the potential impact of drought conditions on final yield.

Those who manage water for multiple uses at regional scales are primarily concerned about future water supplies. These hydrologists are interested in the effects of precipitation (including snowfall) shortfalls on surface or subsurface water supplies, as evidenced by streamflow, reservoir and lake levels, and groundwater table elevations. The frequency and severity of such a 'hydrological' drought is often defined on a watershed or river basin scale. Although all droughts originate with a deficiency of precipitation, hydrologists focus on how this deficiency affects the hydrological system.

Hydrological droughts usually occur some time after precipitation deficiencies show up in soil moisture. It takes longer for streamflows or groundwater and surface reservoir levels to react. It takes even longer for these effects to affect various economic sectors. For example, a precipitation deficiency may result in a rapid depletion of soil moisture that is almost immediately discernible to farmers, yet the impact of this deficiency on reservoir levels may not affect hydroelectric power production or recreational uses for many months. Also, water in lakes, reservoirs, and rivers is often used for multiple and competing purposes (irrigation, recreation, navigation, hydropower, wildlife habitat and so on), further complicating the sequence and quantification of impacts. Competition for water in these storage systems escalates during droughts, and hence conflicts among water users may increase as well.

Developing a history of drought frequency, severity, duration and impacts for a region provides a greater understanding of the region's drought characteristics and the probability of drought recurrence at various levels of severity. Even though these characteristics could be changing along with a changing climate, information of this type is beneficial in the development of drought response and mitigation strategies and preparedness plans. These plans need to be kept current as conditions change.

Figure C.4 summarizes the sequence and consequences of drought events, indicating which of these are of primary interest to meteorologists, agriculturalists and hydrologists.

## 4. Causes of Droughts

Understanding what causes droughts helps us predict them. This section briefly outlines the relation of droughts to normal variations in regional weather, and how weather is related to global climate patterns.

The immediate cause of droughts is the downward movement of air (subsidence). This causes compressional warming or high pressure that inhibits cloud formation and results in lower relative humidity and less precipitation. Most climatic regions experience high-pressure dominance, often depending on the season. Prolonged droughts occur when large-scale high-pressure anomalies in atmospheric circulation patterns persist for months or seasons (or longer). The extreme drought that affected the United States and Canada in 1988–89 resulted from the persistence of a large-scale atmospheric circulation anomaly. Regions under the influence of semi-permanent high pressure during all or a major portion of the year are usually deserts, such as the Sahara and Kalahari deserts of Africa and the Gobi Desert of Asia.

The ability to predict droughts is limited and differs by region, season and climatic regime. It is difficult to predict a drought a month or more in advance for most locations. Predicting droughts depends on the ability to forecast precipitation and temperature, which are inherently variable. Anomalies of precipitation and temperature may last from several months to several decades. How long they last depends on air–sea interactions, soil moisture and land surface processes, topography, internal dynamics, and the accumulated influence of dynamically unstable synoptic weather systems at the global scale.

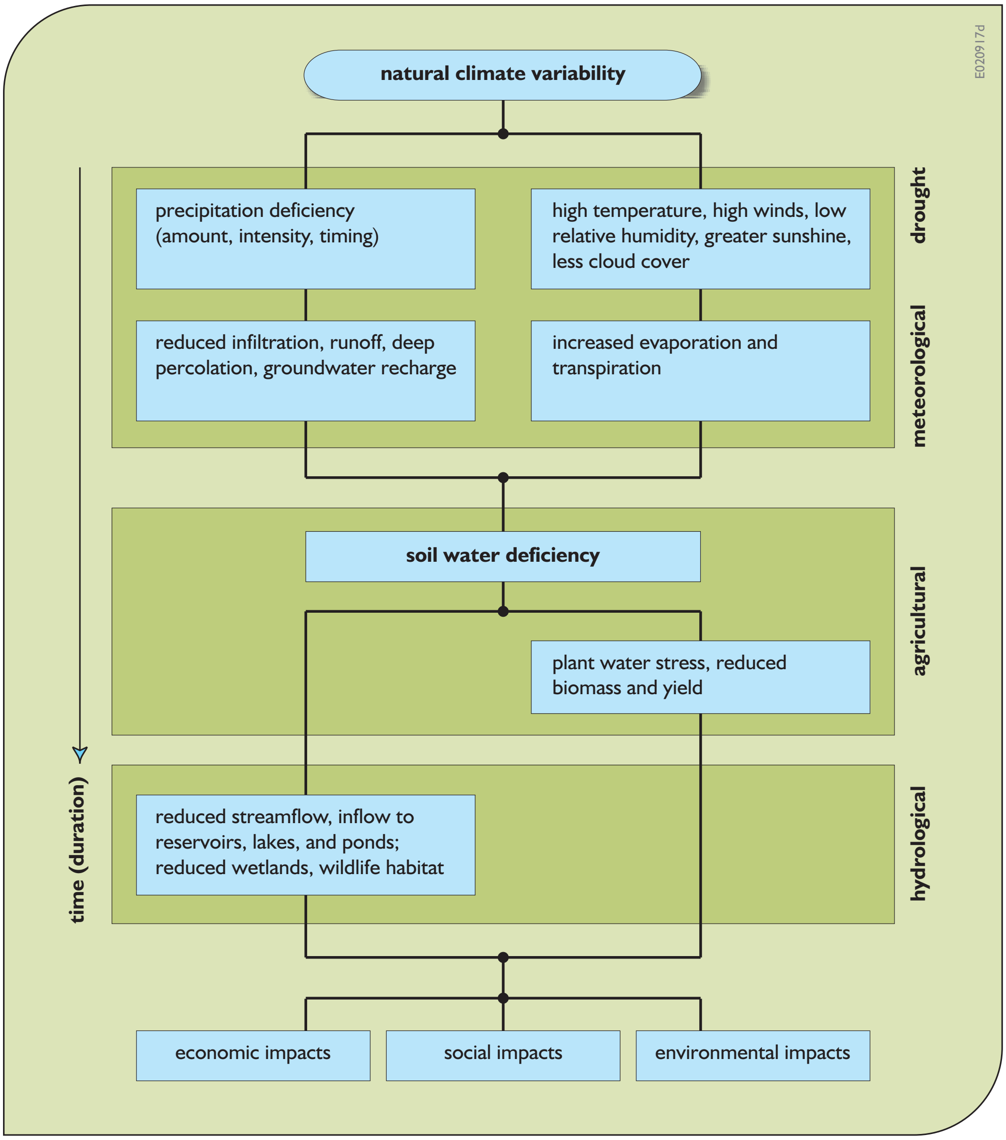



**Figure C.4.** A summary of the sequence and consequences of drought events, and the way meteorologists, agriculturalists and hydrologists view them.

## 4.1. Global Patterns

Global patterns of climatic variability tend to recur periodically with enough frequency and with similar characteristics over a sufficient length of time to offer opportunities for improved long-range climate predictions. One such pattern is called El Niño/Southern Oscillation (ENSO). Off the western coast of South America every two to seven years, ocean currents and winds shift. This brings warm water westward to displace the nutrient-rich cold water that normally wells up from deep in the ocean. The invasion of warm water disrupts both the marine food chain and the economies of coastal communities that depend on fishing and related industries. Because the phenomenon peaks around the Christmas season, the fishermen who first observed it named it El Niño ('the Christ Child' or the little boy).

In recent decades, scientists have recognized that El Niño is linked with other shifts in global weather patterns. The upper half of Figure C.5 illustrates the El Niño conditions compared to normal conditions.

The intensity and duration of El Niño events are varied and hard to predict. They typically last anywhere from fourteen to twenty-two months, but they can be much shorter or longer. El Niño often begins early in the year and peaks between the following November and January,

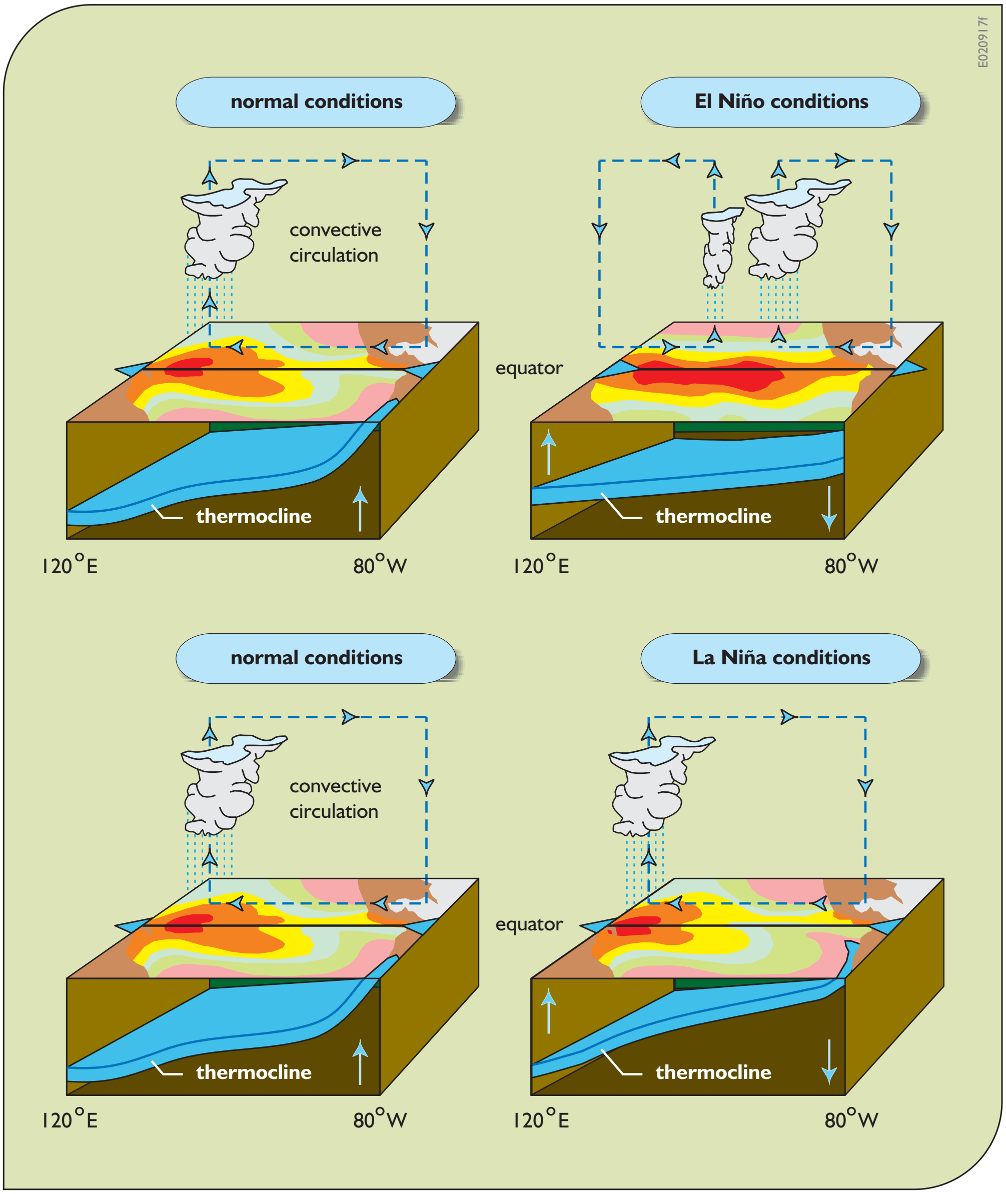



**Figure C.5.** Normal, El Niño, and La Niña conditions in the mid-Pacific Ocean.

but no two events in the past have behaved in the same way.

Often accompanying El Niño is the Southern Oscillation, a 'seesaw of atmospheric pressure between the eastern equatorial Pacific and Indo-Australian areas' (Glantz et al., 1991). The term Southern Oscillation refers to the atmospheric component of the relationship, shown in Figure C.5, and El Niño represents the oceanic property in which sea surface temperatures are the main factor.

During an El Niño/Southern Oscillation event, the Southern Oscillation is reversed. Generally, when pressure is high over the Pacific Ocean, it tends to be low in the eastern Indian Ocean, and vice versa (Maunder, 1992). ENSO events are those in which both a Southern Oscillation extreme and an El Niño occur together.

La Niña is the counterpart of El Niño and represents the other extreme of the ENSO cycle. In this event, the sea surface temperatures in the equatorial Pacific drop well below normal levels and advect to the west, while the trade winds are unusually intense rather than weak. La Niña cold event years often (but not always) follow El Niño warm event years. The lower half of Figure C.5 compares La Niña and normal conditions.

### 4.2. Teleconnections

Shifts in sea surface temperatures off the west coast of South America are just one part of the coupled interactions of atmosphere, oceans and land masses. Statistically, El Niño/Southern Oscillation can account at most for about half of the inter-annual rainfall variance in eastern and southern Africa (Ogallo, 1994). Yet many of the more extreme anomalies, such as severe droughts, flooding and hurricanes, have strong teleconnections to ENSO events. Teleconnections are defined as atmospheric interactions between widely separated regions (Glantz, 1994). Understanding these teleconnections can help in forecasting droughts, floods and tropical storms (hurricanes).

During an ENSO event, drought can occur virtually anywhere in the world, though the strongest connections between ENSO and intense drought are in Australia, India, Indonesia, the Philippines, Brazil, parts of eastern and southern Africa, the western Pacific Basin islands (including Hawaii), central America and various parts of the United States. Drought occurs in each of the above regions at different times (seasons) during an ENSO event and in varying degrees of magnitude.

In the Pacific Basin, Indonesia, Fiji, Micronesia and Hawaii are usually prone to droughts during ENSO events. Virtually all of Australia is subjected to abnormally dry conditions in these periods, but the eastern half has been especially prone to extreme drought. Bush fires and crop failures usually follow. India too has been subjected to drought through a suppression of the summer monsoon season that seems to coincide with ENSO events in many cases. Eastern and southern Africa also show a strong correlation between ENSO events and a lack of rainfall that brings on drought in the Horn region and areas south of there. Central America and the Caribbean islands are also abnormally dry during El Niño events.

ENSO events seem to have a stronger influence on regions in the lower latitudes, especially in the equatorial Pacific and bordering tropical areas. The relationships in the mid-latitudes are less pronounced and less consistent in the way El Niño influences wet or dry weather patterns. The intensity of the anomalies in these regions is also less consistent than those of the lower latitudes.

Incorporating some of these teleconnections within numerical computer models of global climate can improve their predictive capabilities. Predicting when an ENSO event might occur is much easier than predicting its duration and intensity, however. Predictive models are becoming more sophisticated and more effective in many respects due in part to the expanded data sets that are available for the equatorial Pacific region. Some models have predicted ENSO events a year or more in advance.

ENSO forecasts can help countries anticipate and mitigate droughts and floods, and can be useful in agricultural planning. Countries that are in latitudes with strong El Niño connections to weather patterns, such as Brazil, Australia, India, Peru and various African nations, use predictions of near-normal conditions, weak El Niño conditions, strong El Niño conditions or a La Niña condition to help agricultural producers select the crops most likely to be successful in the coming growing season. In countries or regions with a famine early warning system (FEWS) in place, ENSO forecasts can play a key role in mitigating the effects of floods or droughts that can lead to famine. Famine, like drought, is a slow-onset disaster, so forewarning may enable countries to substantially reduce, if not eliminate, its worst impacts.

ENSO advisories are used to a lesser extent in planning in North America and other extratropical countries because the links between ENSO and weather patterns are less clear in these areas. As prediction models improve, the role of ENSO advisories in planning in mid-latitude countries will undoubtedly increase.

### 4.3. Climate Change

Climate change is associated with the greenhouse effect. The greenhouse effect is a naturally occurring phenomenon necessary to sustain life on earth. In a greenhouse, solar radiation passes through a mostly transparent piece of glass or plastic and warms the inside air, surface and plants. As the temperature increases inside the greenhouse, the interior radiates energy back to the outside and eventually a balance is reached.

The earth is in a greenhouse: its atmosphere is the 'piece of glass or plastic'. Short-wave radiation from the sun passes through the earth's atmosphere. Some of this radiation is reflected back into space by the atmosphere, some is absorbed by the atmosphere, and some makes it to the earth's surface. There, it is either reflected or absorbed. The earth, meanwhile, emits long-wave radiation outward. Gases within the atmosphere absorb some of this long-wave radiation and re-radiate it back to the surface. It is because of this greenhouse-like function of the atmosphere that the average global temperature of the earth is what it is, about 15 °C (59 °F). Without the atmosphere and these gases, the average global temperature would be a chilly −18 °C (0 °F). Life that exists on earth today would not be possible without these atmospheric greenhouse gases. They include carbon dioxide ($CO_2$), water vapour ($H_2O$), methane ($CH_4$), nitrous oxide ($N_2O$), chlorofluorocarbons (CFCs), and ozone ($O_3$).

Increasing concentrations of greenhouse gases will increase the temperature of the earth. Currently both are increasing. Carbon dioxide is responsible for approximately half of the greenhouse gas increase. Concentrations of $CO_2$ have been monitored at the Mauna Loa Observatory in Hawaii since 1958. Measurements of $CO_2$ concentrations from before 1958 are based on air bubbles within ice cores taken from around the world. In the middle of the eighteenth century, levels of $CO_2$ were around 270 parts per million (ppm). By the end of the twentieth century these levels were about 370 ppm. They are currently increasing at about 0.4% per year (IPCC, 2001). About 70–90% of the $CO_2$ added to the atmosphere is due to the burning of fossil fuels, and much of the rest is from deforestation (OTA, 1993). Methane, $N_2O$, and CFCs are increasing at similar rates. At these rates the concentration of $CO_2$ will double pre-industrial concentrations by about 2100 (IPCC, 2001).

Scientists incorporate this information within large-scale models of the atmosphere called general circulation models (also called global climate models, or GCMs). These models are designed to simulate global atmospheric conditions and make projections of the future climate. Although there are differences among the GCM projections, the models are in general agreement that, as a result of increasing greenhouse gas concentrations, the average global temperature will increase by about 1.4–5.8 °C (2.52–10.44 °F) by 2100 (IPCC, 2001). In the past century, the global average surface temperature has increased 0.60 °C (1.08 °F). This increase by itself is within the normal variability and, although it may be a result of climate change, it cannot be used as definitive proof that recent human activities have contributed noticeably to global warming. Between 1860 and 2000, however, the nine warmest years for the global temperature occurred after 1980 (IPCC, 2001).

With the projected global temperature increase, the global hydrological cycle could also intensify. GCMs indicate that global precipitation could increase 7–15%. Meanwhile, global evapotranspiration could increase 5–10% (OTA, 1993). The combined impacts of increased temperature, precipitation, and evapotranspiration would affect snowmelt, runoff, and soil moisture conditions. The global climate models (GCM) generally show that precipitation will increase at high latitudes and decrease at low and mid-latitudes. If this happens, evapotranspiration in mid-continental regions will be greater than precipitation. This could lead to more severe, longer-lasting droughts in these areas. In addition, the increased temperatures alone will expand ocean waters and this will lead to an estimated sea level rise of about half a centimetre per year (OTA, 1993).

One of the weaknesses of GCM climate change predictions is that they cannot adequately take into account factors that might influence regional and river basin climates, such as the local effects of mountains, coastlines, large lakes, vegetation boundaries and heterogeneous soils, or the ways that human activities affect those predictions. Thus, it is difficult to transfer GCM predictions to regional and basin scales. GCMs also cannot predict changes in the frequency of extreme events. Some believe the occurrence of extreme events such as droughts is increasing. Although the losses due to these events have increased worldwide, this is often the result of increased vulnerability rather than an increased number of events.

No one knows who will lose or who will benefit from a $CO_2$-induced climate change. Clearly, coastal regions will be likely to experience a slow rise in average sea levels. But some interior regions might benefit from gains in agricultural production resulting from the indirect effects of a warmer climate and adequate precipitation, especially in higher latitudes across Canada and Russia. The increased $CO_2$ might also directly increase plant growth and productivity as well. This is known as the $CO_2$ fertilization effect. Laboratory experiments have shown that increased $CO_2$ concentrations potentially promote plant growth and ecosystem productivity by increasing the rate of photosynthesis, improving nutrient uptake and use, increasing water-use efficiency, and decreasing respiration, along with several other factors (OTA, 1993). Increased ecosystem productivity could draw $CO_2$ from the atmosphere, thereby diminishing concerns about global warming (OTA, 1993). Whether any benefits would result from the $CO_2$ fertilization effect within the complex interactions of natural ecosystems is still unknown, however.

### 4.4. Land Use

Although climate is a primary contributor to droughts, other factors such as changes in land use (e.g. deforestation), land degradation and the construction of dams all affect the hydrological characteristics of a basin. Changes in land use upstream may reduce infiltration and thereby increase runoff rates, resulting in more variable streamflow, less groundwater retention and a higher incidence of hydrological drought downstream. Bangladesh, for example, has shown an increased frequency of water shortages in recent years because of land use changes that have occurred within its territory and in neighbouring countries.

Land use change is one of the ways human actions can alter the frequency of water shortage even when no statistical change in the frequency or intensity of precipitation has been observed.

## 5. Drought Indices

Drought indices combine various indicators of drought into a single index to help in identifying the onset as well as the severity of a drought. Different indices have been proposed and are used for different purposes. This section identifies a few of the more commonly used indices. No index is perfect for all situations. Hence, most water supply planners find it useful to track more than one index before making decisions that will affect people's welfare.

### 5.1. Percent of Normal Indices

The *percent of normal* indices are easily calculated and well suited to the needs of weathercasters and the general public. They can be associated with any hydrological variable, such as precipitation, soil moisture, groundwater level or reservoir storage volume. The percent of normal index based on precipitation is one of the simplest measurements of rainfall or snowfall for a location. Such analyses are effective when used for a single region or a single season. The index value is 100 times the actual value divided by the average value (typically considered to be a thirty-year mean). Normal for a specific location is considered to be 100%.

This index can be calculated for a variety of time scales. Usually these range from a single month to a group of months representing a particular season, to an annual or water year.

One potential disadvantage of using the percent of normal value is that the mean, or average, is often not the same as the median (the value exceeded 50% of the time in a long-term climate record). This is because most random hydrological values on monthly or seasonal scales are not normally distributed. Some people assume the use of the percent of normal comparison is based on median rather than mean values, and this can cause confusion. The long-term precipitation record for the month of January in Melbourne, Australia, can serve to illustrate the potential for misunderstanding. The median January precipitation is 36.0 mm (1.4 in.), meaning that in half the January months less than 36.0 mm has been recorded, and in the other half of the January months more than 36.0 mm has been recorded. However, a monthly January total of 36.0 mm would be only 75% of the historical mean, which is often considered to be quite dry (Willeke et al., 1994).

### 5.2. Standardized Precipitation Index

The realization that a precipitation deficit has different impacts on groundwater, reservoir storage, soil moisture, snow pack and streamflow led McKee et al. (1993) to develop the Standardized Precipitation Index (SPI). This index quantifies the precipitation deficit for multiple time scales that reflect the impact of drought on the availability of different sources of water. Soil moisture conditions respond to precipitation anomalies on a relatively short time scale. Groundwater, streamflow and reservoir storage reflect the longer-term precipitation anomalies. For these reasons, McKee et al. (1993) originally calculated the SPI for three, six, twelve, twenty-four, and forty-eight-month time scales.

The SPI calculation for any location is based on the long-term precipitation record fitted to a probability distribution. This distribution is then transformed into a normal distribution so that the mean SPI for the location and desired period is zero (Edwards and McKee, 1997). Positive SPI values indicate greater than median or mean precipitation, and negative values indicate less than median or mean precipitation. Because the SPI is normalized, wetter and drier climates can be represented in the same way, so wet periods can also be monitored using the SPI.

McKee et al. (1993) used the classification system shown in Table C.2 to define drought intensities resulting from the SPI. They also defined the criteria for a drought event for any of the time scales. A drought event occurs any time the SPI is continuously negative and reaches an intensity of −1.0 or less. The event ends when the SPI becomes positive. Each drought event, therefore, has a duration defined by its beginning and end, and an intensity for each month of the drought event. The absolute sum of the negative SPI for all the months within a drought event can be termed the drought's 'magnitude'.

Using an analysis of stations across Colorado in the western United States, McKee determined that, as defined by the SPI, western Colorado is in mild drought 24% of the time, in moderate drought 9.2% of the time, in severe drought 4.4% of the time, and in extreme drought 2.3% of the time (McKee et al., 1993). Because the SPI is standardized, these percentages are expected from a normal distribution of the SPI. The 2.3% of SPI values within the 'extreme drought' category is a percentage that is typically expected for an 'extreme' event (Wilhite, 1995). In contrast, the Palmer index, to be discussed next, reaches its 'extreme' category more than 10% of the time across portions of the central Great Plains. This standardization allows the SPI to determine the rarity of a current drought, as well as the probability of the precipitation necessary to end the current drought (McKee et al., 1993).

| SPI values | |
|---|---|
| 2.0+ | extremely wet |
| 1.5 to 1.99 | very wet |
| 1.0 to 1.49 | moderately wet |
| -0.99 to 0.99 | near normal |
| -1.00 to -1.49 | moderately dry |
| -1.50 to -1.99 | severely dry |
| -2.00 and less | extremely dry |

E021014b

**Table C.2. Standardized precipitation index.**

## 5.3. Palmer Drought Severity Index

The Palmer drought severity index (PDSI) is a soil moisture index designed for relatively homogeneous regions. Many US government and state agencies rely on it to trigger drought relief programmes. The index measures the departure of the soil moisture supply from demand (Palmer, 1965). The objective of the PDSI is to provide standardized measurements of moisture conditions so that comparisons using the index can be made between locations and between months. Complete descriptions of the equations can be found in the original study by Palmer (1965) and in the analysis by Alley (1984).

The PDSI responds to weather conditions that have been abnormally dry (or abnormally wet). It is calculated with the use of precipitation and temperature data, as well as the local available water content (AWC) of the soil. It does not consider streamflow, lake and reservoir levels, or other longer-term hydrological variable values that indeed may still show a drought (Karl and Knight, 1985). Human impacts on the water balance, such as irrigation, are also not considered.

Palmer developed the PDSI to include the duration of a dry or wet spell period. His motivation was as follows: an abnormally wet month in the middle of a long-term drought should not have a major impact on the index, and a series of months with near-normal precipitation following a serious drought does not mean that the drought is over. Therefore, Palmer developed criteria for determining when a drought or a wet spell begins and ends. Palmer (1965) described this effort and gave examples. It is also described in detail by Alley (1984).

In near-real time, Palmer's index is no longer a meteorological index but becomes a hydrological one, referred to as the Palmer hydrological drought index (PHDI). It is based on moisture inflow (precipitation), outflow and storage, and does not take into account the long-term trend (Karl and Knight, 1985).

A modified method to compute the PDSI was proposed by Heddinghaus and Sabol, 1991. This modified PDSI differs from the PDSI during transition periods between dry and wet spells. Because of the similarities between these Palmer indices, the terms Palmer index and Palmer drought index have been used to describe general characteristics of the indices.

The Palmer index varies between roughly −6.0 and +6.0. Palmer arbitrarily selected the classification scale of moisture conditions, shown in Table C.3, on the basis of his original study areas in the central United States (Palmer, 1965). Ideally, the Palmer index is designed so that a −4.0 in one region has the same meaning in terms of the moisture departure from a climatological normal as a −4.0 in another region (Alley, 1984). The Palmer index has typically been calculated on a monthly basis, and a long-term archive (from 1895 to present) of the monthly PDSI values for every climate division in the United States exists at the National Climatic Data Center. In addition, weekly Palmer index values (modified PDSI values) are calculated and mapped (Figure C.6) for the climate divisions during every growing season.

The Palmer index has been widely used for a variety of applications across the United States. It is most effective at measuring impacts sensitive to soil moisture conditions, such as agriculture (Willeke et al., 1994). It has also been useful as a drought-monitoring tool and has been used to trigger actions associated with drought contingency plans (Willeke et al., 1994). Alley (1984) identified three positive characteristics of the Palmer index that contribute to its popularity: it provides decision-makers with a measurement of the abnormality of recent weather for a region; it provides an opportunity to place current conditions in historical perspective; and it provides spatial and temporal representations of historical droughts. Several states in the USA, including New York, Colorado, Idaho and Utah, use the Palmer index as part of their drought monitoring systems.

The Palmer index has some limitations, and these are described in detail by Alley (1984) and Karl and Knight (1985). Also, while the index has been applied within the United States it has little acceptance elsewhere (Kogan, 1995; Smith et al., 1993). It does not do well in regions where there are extremes in the variability of rainfall or runoff, such as Australia and South Africa. Another weakness in the Palmer index is that the 'extreme' and 'severe' classifications of drought occur with a greater frequency in some regions than in others (Willeke et al., 1994). This limits the accuracy of comparisons of the intensity of droughts between two regions and makes planning response actions based on a given intensity more difficult.

E021014c

**Palmer classifications**

| | |
|---|---|
| 4.0 or more | extremely wet |
| 3.0 to 3.99 | very wet |
| 2.0 to 2.99 | moderately wet |
| 1.0 to 1.99 | slightly wet |
| 0.5 to 0.99 | incipient wet spell |
| 0.49 to -0.49 | near normal |
| -0.5 to -0.99 | incipient dry spell |
| -1.9 to -1.99 | mild drought |
| -2.0 to -2.99 | moderate drought |
| -3.0 to -3.99 | severe drought |
| -4.0 or less | extreme drought |

**Table C.3.** The Palmer drought severity index.

## 5.4. Crop Moisture Index

Whereas the PDSI monitors long-term meteorological wet and dry spells, the crop moisture index (CMI) was designed to evaluate short-term moisture conditions across major crop-producing regions. It is not intended to assess long-term droughts. It reflects moisture supply in the short term across major crop-producing regions.

The CMI uses a meteorological approach to monitor week-to-week crop conditions. It is based on the mean temperature and total precipitation for each week within a climate division, as well as the CMI value from the previous week. The index responds rapidly to changing conditions, and is weighted by location and time so that maps, which commonly display the weekly CMI across a region, can be used to compare moisture conditions at different locations.

## 5.5. Surface Water Supply Index

The surface water supply index (SWSI, pronounced 'swa zee') is designed to complement the Palmer where mountain snow pack is a key element of water supply. It

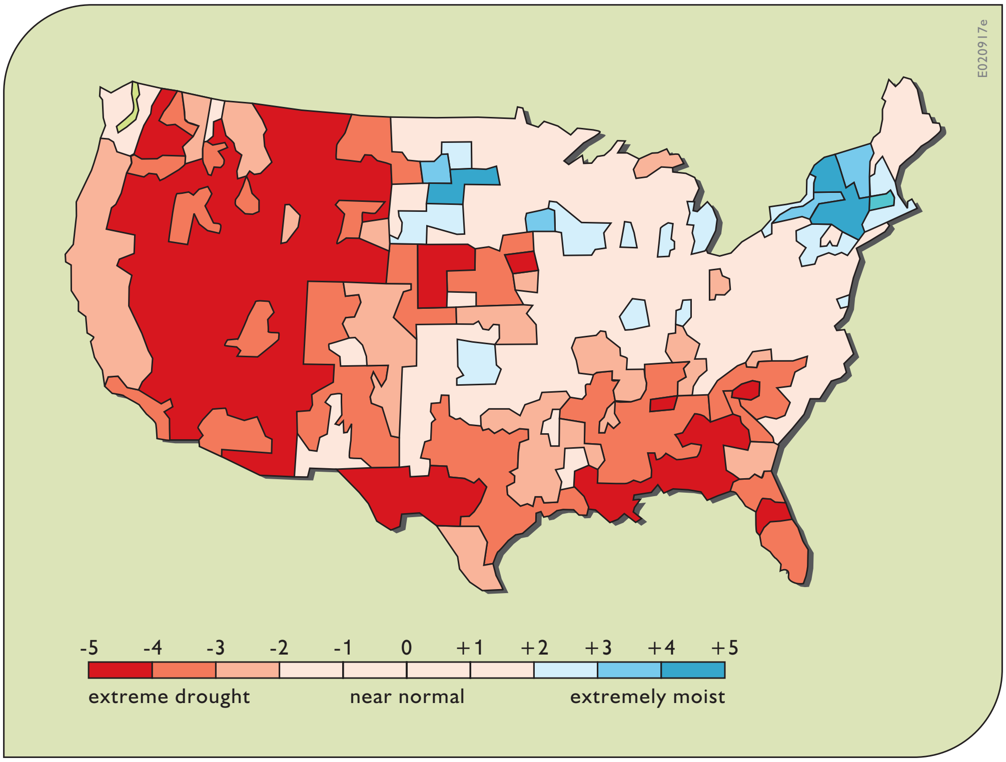



**Figure C.6.** Palmer index values for the United States during a week in August, 2000.

is calculated for river basins on the basis of snow pack, streamflow, precipitation and reservoir storage (Shafer and Dezman, 1982; Dezman et al., 1982). The index is unique to each basin, which limits interbasin comparisons.

Shafer and Dezman (1982) designed the SWSI to be an indicator of the conditions of surface water in which mountain snow pack is a major component. The objective of the SWSI was to incorporate both hydrological and climatological features into a single index value resembling the Palmer index for each major river basin in the state of Colorado. These values could be standardized to allow comparisons between basins. Four inputs are required within the SWSI: snow pack, streamflow, precipitation and reservoir storage. Because it is dependent on the season, the SWSI is computed with only snow pack, precipitation and reservoir storage in the winter. During the summer months, streamflow replaces snow pack as a component within the SWSI equation.

To determine the SWSI for a particular basin, monthly data are collected and summed for all the precipitation stations, reservoirs and snow pack/streamflow measuring stations over the basin. Each summed component is normalized using a frequency analysis gathered from a long-term data set. The probability of non-exceedance (the probability that subsequent sums of that component will exceed the current sum) is determined for each component on the basis of the frequency analysis. This allows comparisons of the probabilities to be made between the components. Each component has a weight assigned to it depending on its typical contribution to the surface water within that basin. These weighted components are summed to determine a SWSI value representing the entire basin. Like the Palmer index, the SWSI is centred on zero and has a range between −4.2 and +4.2.

## 5.6. Reclamation Drought Index

Like the SWSI, the reclamation drought index (RDI) is calculated at the river basin level, incorporating temperature as well as precipitation, snow pack, streamflow and reservoir levels as input. By including a temperature component, it also accounts for evaporation. It is used by the US Bureau of Reclamation as a trigger to release drought emergency relief funds. The RDI classifications are listed in Table C.4.

The RDI was developed as a tool for defining drought severity and duration, and for predicting the onset and end of periods of drought. The impetus to devise it came

| RDI classifications | | | |
|---|---|---|---|
| 4.0 | or more | | extremely wet |
| 1.5 | to | 4.0 | moderately wet |
| 1 | to | 1.5 | normal to mild wetness |
| 0 | to | 1.5 | normal to mild drought |
| -1.5 | to | -4.0 | moderate drought |
| -4.0 | or less | | extreme drought |

E021014d

Table C.4. RDI classifications.

| decile classifications | | |
|---|---|---|
| deciles 1 - 2: | lowest 20 % | much below normal |
| deciles 3 - 4: | lowest next 20 % | below normal |
| deciles 5 - 6: | middle 20 % | near normal |
| deciles 7 - 8: | highest next 20 % | above normal |
| deciles 9 - 10: | highest 20 % | much above normal |

E030526o

Table C.5. Decile classifications.

from the US Reclamation States Drought Assistance Act of 1988, which allows states to seek assistance from the US Bureau of Reclamation to mitigate the effects of drought.

The RDI differs from the SWSI in that it builds a temperature-based demand component and duration into the index. The RDI's main strength is its ability to account for both climate and water supply factors.

## 5.7. Deciles

This index groups monthly precipitation occurrences into deciles so that, by definition, 'much lower than normal' weather cannot occur more often than 20% of the time. These deciles are shown in Table C.5.

Arranging monthly precipitation data into deciles is another drought-monitoring technique. Developed by Gibbs and Maher (1967), it removes some of the weaknesses in the 'percent of normal' approach. The technique divides the distribution of occurrences over a long-term precipitation record into tenths of the distribution. Each of these categories is a decile. The first decile is the rainfall amount not exceeded by the lowest 10% of the precipitation occurrences. The second decile is the precipitation amount not exceeded by the lowest 20% of occurrences. These deciles continue until the rainfall amount identified by the tenth decile is the largest precipitation amount within the long-term record. By definition, the fifth decile is the median; it is the precipitation amount not exceeded by 50% of the occurrences over the period of record. The deciles are grouped into five classifications, as shown in Table C.5.

The decile method was selected as the meteorological measurement of drought within the Australian Drought Watch System because it is relatively simple to calculate and requires fewer data and assumptions than the Palmer Drought Severity Index (Smith et al., 1993). In this system, farmers and ranchers can only request government assistance if the drought is shown to be an event that occurs only once in twenty or twenty-five years (deciles 1 and 2 over a 100-year record) and has lasted longer than twelve months (White and O'Meagher, 1995). This uniformity in drought classifications, unlike a system based on the percent of normal precipitation, has assisted Australian authorities in determining appropriate drought mitigation responses.

## 5.8. Method of Truncation

The method of truncation uses historic records of streamflow, precipitation, groundwater drawdown, lake elevation and temperature (Chang and Kleopa, 1991). The historic data are sorted in ascending order, and trigger levels are determined from the truncation level specified (Figure C.7).

For example, if stage 1 drought is defined as the 70% level, this corresponds to streamflows that are less than 70% of all flows. When using this method, drought events of higher severity are 'nested' inside drought events of lower severity: that is to say, a 90% drought implies the occurrence of a 70% and 80% drought (if those are the trigger levels selected). The highest level of severity determines the decisions to make, as a water management system must deal with the most severe shortage that occurs.

## 5.9. Water Availability Index

The water availability index (WAI) relates current water availability to historical availability during periods of drought by measuring the deviation-from-normal rainfall

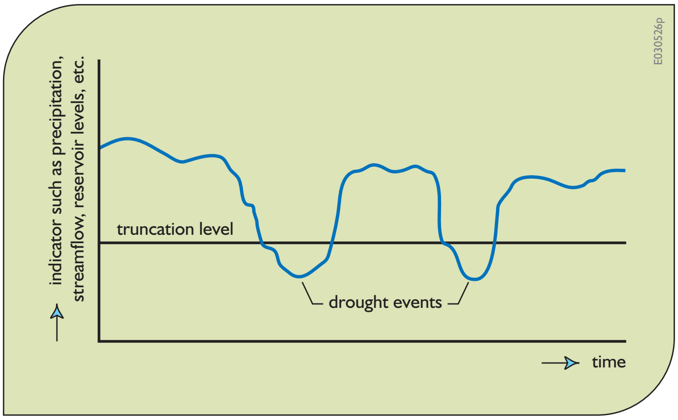



**Figure C.7. Method of truncation to define drought conditions (Fisher and Palmer, 1997).**

| week | beginning storage volume | inflow volume per week | demand volume per week | ending storage volume | demand met? | DSR days |
|---|---|---|---|---|---|---|
| begin | 100 | 40 | 40 | 100 | | 7 |
| + 1 | 100 | 60 | 40 | 120 | | 14 |
| + 2 | 120 | 20 | 40 | 100 | | 21 |
| + 3 | 100 | 20 | 40 | 80 | | 28 |
| + 4 | 80 | 0 | 40 | 40 | | 35 |
| + 5 | 40 | 20 | 40 | 20 | | 42 |
| + 6 | 20 | 10 | 40 | -10 | no | |
| + 7 | -10 | 60 | 40 | 10 | | |

E030526q

**Table C.6. Calculation of DSR index value based on weekly time steps and forecast inflow and demands (from Fisher and Palmer, 1997).**

over the prior four months (Davis and Holler, 1987). The WAI is multiplied by a constant to make the index fall between 0 and 10, with zero representing normal conditions and ten a severe drought. Hoke (1989) included a factor for minimum and maximum desirable reservoir pool elevation, and Raney (1991) included volume and drainage area factors to account for multiple reservoirs within one watershed.

## 5.10. Days of Supply Remaining

The days of supply remaining (DSR) index includes current reservoir storage, forecast future inflows from precipitation and snowmelt, and predicted demands for municipal supply and instream flows (Fisher and Palmer, 1997). It provides an objective and easily understood measure of the supply of water. As new sources of supply or new demands occur, the index remains appropriate.

DSR is calculated by predicting future inflows and demands and determining when the inflows and existing supply will no longer be adequate to meet demands. A simplified example shown in Table C.6 illustrates this calculation.

The DSR values are calculated at the beginning of each time step, for example, a week. At that time, forecasts are made of the subsequent weeks' inflows and demands. Inflows are added and demands subtracted for each successive week until the remaining supply is less than the demand.

In the example shown in Table C.6, there are 100 units of storage currently available. Forecasts show that the projected demand cannot be met after six weeks from now. There is adequate supply for the current week plus the following five weeks. Thus, only forty-two days of supply remains. A water management agency presented with the data in this example could implement restrictions to try to reduce demand by 5%. If successful that would maintain adequate levels of storage until the expected reservoir refill begins in seven weeks. Alternatively, a new supply source could be sought, or other demand and supply management options could be analysed. Determining just what actions one should take (given the uncertainty of supply and demand forecasts) can be based, in part, on such analyses, including simulations of alternative actions given supply and demand situations experienced in the past.

Palmer et al. (2002) describe doing this for the City of Atlanta in southeastern United States. Determining in advance what actions are appropriate under what specific conditions or values of DSR and getting them accepted makes drought management substantially easier when under the stress of an actual drought.

The particular values of DSR or any other index value that trigger management decisions are called drought triggers.

## 6. Drought Triggers

Drought management is tricky. Without knowing just how severe a drought may become, it is difficult to know what actions are best to take, if any, when a drought is likely to have begun. Water managers must take appropriate action to reduce future losses and not increase them by imposing unnecessary restrictions when they are not needed (Wright et al., 1986). Taking actions that impose hardships on the public and the water utility when it later turns out they were unnecessary is not likely to engender compliance with water saving actions the next time they are ordered. One way to make drought management easier is to establish a sequence of increasingly strict conservation measures based on a sequence of drought triggers, and seek the public's approval.

A drought trigger is the specific value of a drought indicator that activates a management response. For example, a drought trigger could be a reservoir decreasing below 50% of its active storage capacity. In a drought contingency plan, trigger levels can be varied to alter the sensitivity of the response and the effectiveness of the plan. Defining drought triggers can be difficult. Trigger levels change over time: that is, an appropriate trigger level for a particular system may change dramatically if that system's infrastructure is expanded or if water demands change dramatically. Urban water triggers are often quite different from agricultural drought triggers, as the urban infrastructure can often mitigate the impacts of short-term droughts.

Short-term responses may include the initiation of outdoor water use bans, an increase in the price of water or the use of printed media to inform the public of water supply problems. Drought management plans for many urban areas are often developed with four to five levels of responses, all of which encourage different levels of demand reduction or supply augmentation. The appropriateness or effectiveness of particular drought measures is dependent upon the type of user. An outdoor water ban, for instance, may be effective for a residential community but not for a heavy industrial complex.

Specific values of any of the drought indices just reviewed in the previous section could be used for a drought trigger. Since each has its limitations, triggers based on multiple indices appropriate for the region will probably be the most reliable and robust (Alley, 1985; Dracup et al., 1980; Johnson and Kohne, 1993; Mather, 1985; Titlow, 1987).

Numerous utilities and government organizations in many countries have explored the use of triggers. Those using triggers to support decision-making in the United States include the Massachusetts Water Resources Authority (which serves the Boston area), Northern New Jersey, the Interstate Commission on the Potomac River Basin (which facilitates supply operations in Washington, D.C., northern Virginia and southern Maryland), the Bureau of Water Works for the City of Portland, and the Seattle Water Department and Denver Water (Fisher and Palmer, 1997; Hrezo et al., 1986; Nault et al., 1990; USACE, 1991).

Deciding on what the trigger indices should be, what the trigger values of those indices should be, and what decisions to make when those trigger index values have been reached is not a trivial exercise. Performing simulations of alternative actions given supply and demand situations experienced in the past can help identify the possible range of likely impacts associated with any particular set of trigger values and associated decisions (Palmer et al., 2002). Performing these analyses in a real-time virtual drought simulation exercise involving all relevant agencies and decision-makers may be even more effective.

## 7. Virtual Drought Exercises

Drought management requires more than just a plan. Its implementation needs public support and financial resources. The plan needs to be kept up to date. People who are going to be implementing it need to be ready to do that when a drought comes. Virtual drought exercises are one way to accomplish and maintain a state of readiness.

A virtual drought is a fire drill, a test during which water managers, public health officials and other stakeholders put a drought management plan into practice using a hypothetical, computer-generated drought event. A good virtual drought exercise will raise important issues and provide experience that can be applied in real droughts. It will also keep those from different agencies who must work closely together during times of droughts in closer contact with each other.

For over twenty years, the Interstate Commission on the Potomac River Basin (ICPRB) has been conducting

**Box C.1. A Recent History of the Potomac River**

In 1963, the US Army Corps of Engineers conducted a comprehensive study of the Potomac River basin to identify solutions to the anticipated demand deficits during low river flows. They proposed the construction of sixteen large multipurpose reservoirs in the Potomac River basin. This proposal was not accepted by the public, nor by Congress who would pay the bill. Flow shortages in which unrestricted water use would cause the river to run dry, coupled with increasing population in the 1960s and 1970s, underlined the need to continue to look for solutions. One of the sixteen proposed reservoir projects was built: Jennings Randolph Lake (originally called Bloomington Lake).

Other structural solutions were examined. Interbasin transfers were analysed. A pilot treatment plant and an emergency pumping station were constructed.

Concurrently, an academic study was being conducted that combined distribution areas of the three major Washington metropolitan area utilities into a single regional demand centre. Through simulation modelling, the study investigated the benefits derived from the coordinated operation of all the resources then available to all three utilities. The study showed that coordinated management of the water resources from a total integrated systems perspective led to gains in the reliability of the water resource. The results of the latter analysis and its lower cost non-structural features led to the adoption of its results with the signing of the Water Supply Coordination Agreement in 1982.

The management objectives embodied in the agreement are to keep the reservoir storage volumes balanced while meeting environmental requirements and municipal demands for water. The coordinated operations of each utility are expected to meet the expected demands without imposing restrictions through the year 2015 even under a repeat of the drought of record. This is possible because of synergistic gains in total yield realized under the cooperative management strategies. Each utility gives up a small measure of autonomy in order to gain the substantial benefits of reduced capital costs through coordinated cooperative operations of its individually and jointly owned resources. A committee of the water utilities and the Interstate Commission on the Potomac River Basin oversee the coordinated operations.

annual virtual drought exercises for the Washington metropolitan area. During a drought, the Commission coordinates water supply operations of three major water utilities in Washington, D.C., and the adjacent suburbs in the states of Maryland and Virginia. In the drought exercises, the Commission and the utilities practise the operations of the system as they would during an actual drought. Box C.1 gives a history of how this began.

Virtual drought exercises typically have three stages: briefing, gaming and debriefing. During the briefing, the objectives of the exercise are stated, the rules of the exercise are described, and the initial conditions of the simulated drought scenario are introduced. During the gaming, participants discuss management strategies and negotiate decisions to respond to the simulated drought conditions as they progress. Finally, during debriefing, the outcomes of the exercise are evaluated and drought management plans are modified if and as appropriate.

The briefing portion of the exercise provides a foundation for the day, establishing a common understanding of the objectives and how the gaming will take place. Briefing materials are distributed in advance of the exercise to give participants time to clarify questions or misunderstandings. Typically the briefing will:

- define the objectives of the exercise
- introduce participants
- review the roles of participants
- clarify the format of the exercise
- specify the rules for decision-making
- explain the background information
- demonstrate the virtual drought process.

During the gaming portion of the exercise, the participants manage the water resources system during a drought scenario. This requires discussion among the participants and joint decision-making. Guidelines are provided to define options and modes of decision-making, but the negotiation and acceptance of decisions is one of the by-products of this drought rehearsal. Typical responses include water use restrictions, modifying instream flow requirements, revising water pricing, emergency source acquisitions, modifying surface and groundwater system operations, public involvement and regional coordination measures.

During the gaming, the drought event unfolds over time. The participants intervene with actions when necessary to manage the drought. Information to support and/or influence decision-making is provided throughout the exercise. Such information can include the existing drought plan, forecasts of future conditions, system status reports, results of technical analyses, media news briefs and legal actions.

Debriefing is the reflective and evaluative portion of the exercise. The participants' actions and perceptions are discussed and translated into 'lessons learned'. Specific insights on how the participants managed the drought are brought out and recommendations are made. A facilitated group discussion is the best form of debriefing, providing the participants an opportunity for self-observation.

There are clear links between the general drought planning and preparedness process and virtual drought exercises. Virtual drought exercises test a group's abilities to resolve conflicts and make decisions; they allow participants to experience a drought event and practise its management without the dangers of a real event, and they provide an effective means of evaluating a drought plan's likely effectiveness.

The more realistic the virtual drought exercise is, the more beneficial it will be. It is important that those individuals who are responsible for drought management participate, that the participants represent a wide range of perspectives, that the objectives of the exercise are clear to all, and that the exercise be made as realistic as possible. This makes it possible to capture a true sense of how the resource is likely to be managed during the stress of an actual drought.

Virtual drought exercises should be viewed as natural complements to traditional drought planning activities. If properly designed and implemented, these exercises can keep plans up to date, stakeholders informed and water managers' skills current.

## 8. Conclusion

Droughts are unusually severe water shortages. We cannot predict when they will occur or how long they will last, but they will happen and when they do the economic as well as social costs can be substantial. Being prepared for droughts can help to mitigate some of those costs and damages. It costs money to plan, but planning to be prepared for droughts in advance of when they occur saves much more that it costs.

Society's vulnerability to droughts is affected by population density and growth (especially in urban regions), changes in water use trends, government policy, social behaviour, economic conditions, and environmental and ecological objectives. Changes in all of these factors tend to increase the demand for water, and hence society's vulnerability to droughts.

Although droughts are natural hazards, society can reduce its vulnerability and therefore lessen the risks associated with drought events. The impacts of droughts, like those of other natural hazards, can be reduced through planning and preparedness. Drought management is decision-making under uncertainty. It is risk management. Planning ahead to identify effective ways of mitigating drought losses gives decision-makers the chance to reduce both suffering and expense. Reacting without a plan to emergencies in a 'crisis mode' during an actual drought generally decreases self-reliance and increases dependence on government services and donors.

Planning for droughts is essential, but it may not come easily. There are many constraints on planning. For example, it is hard for politicians and the public to be concerned about a drought when they are coping with a flood – or any other more immediate crisis. Unless there is a drought emergency, it is often hard to get support for drought planning. There are always more urgent needs for money and people's attention. Where coordination among multiple agencies can yield real benefits, it is not easy to get it to happen until it becomes obviously essential, for example during a severe drought. Multi-agency cooperation and coordination must be planned for, and perhaps practised in virtual drought management exercises, in advance of the drought. Getting multiple agencies to work together only in a crisis mode is never efficient. Crisis-oriented drought response efforts have been largely ineffective, poorly coordinated, untimely, and inefficient in terms of the resources allocated.

Drought planning will vary from one city or region to another, simply because resources, institutions and populations differ. Although drought contingency plans may vary in detail, they all should specify a sequence of increasingly stringent steps to either augment supplies or reduce demand as the drought becomes more severe – that is, as the water shortage increases. This should happen in such a way as to minimize the adverse impacts of water shortages on public health, consumer activities, recreation, economic activity and the environment in the most cost-effective manner possible.

Drought plans provide a consistent framework to prepare for and respond to drought events. The plans should include drought indicators, triggers, and responses. They should also include provisions for forecasting drought conditions, monitoring and enforcement (Werick and Whipple, 1994). Drought plans should consider a wide range of issues and be compatible with the political and social environments that can affect just what measures can be implemented.

Developing a drought plan and keeping it current is a continuing process that should include an informed public. Drought plans should also include measures to educate the public and keep them aware of the potential risks of droughts, and the measures that will be implemented to mitigate those risks. A comprehensive public information programme should be implemented to achieve public acceptance of and compliance with the plan. At the same time, enforcement measures are often necessary to encourage the public to abide by the water-use restrictions. Enforcement measures traditionally include penalties for noncompliance. They can also include economic incentives such as rebates on low flow showerheads and faucets, and cheaper water rate charges for lower consumption rates.

## 9. References

ALLEY, W.M. 1984. The Palmer drought severity index: limitations and assumptions. *Journal of Climate and Applied Meteorology*, No. 23, pp. 1100–9.

ALLEY, W.M. 1985. The Palmer drought severity index as a measure of hydrological drought. *Water Resources Bulletin*, Vol. 21, No. 1, pp. 105–14.

CHANG, T.J. and KLEOPA, X.A. 1991. A proposed method for drought monitoring. *Water Resources Bulletin*, Vol. 27, No. 2, pp. 275–81.

DAVIS, P.C. and HOLLER, A.G. 1987. Southeastern drought of 1986: lessons learned. In: A.D. Feldman (ed.), *Proceedings of the Engineering Hydrology Symposium*. Williamsburg, Va., 3–7 August 1987. New York, ASCE, pp. 616–21.

DEZMAN, L.E.; SHAFER, B.A.; SIMPSON, H.D. and DANIELSON, J.A. 1982. Development of a surface water supply index: a drought severity indicator for Colorado, In: A.I. Johnson and R.A. Clark (eds.). *Proceedings of the International Symposium on Hydrometeorology*, 13–17 June 1982, Denver, Colorado. Bethesda, Md., American Water Resources Association, pp. 337–41.

DRACUP, J.A.; LEE, K.S. and PAULSON, E.G. 1980. On the definition of droughts. *Water Resources Research*, Vol. 16, No. 2, pp. 297–302.

EDWARDS, D.C. and MCKEE, T.B. 1997. *Characteristics of twentieth century drought in the United States at multiple time scales*. Fort Collins, Colo., Colorado State University, Department of Atmospheric Sciences (Climatology Report Number 97–2).

FISHER, S.M. and PALMER, R.N. 1995. Managing water supplies during drought: the search for triggers. In: M.F. Domenica (ed.), *Proceedings of the 22nd Annual National Conference*, 7–11 May 1995, Cambridge, Mass. Cambridge, Mass., Water Resources Planning and Management Division of ASCE, pp. 1001–4.

GIBBS, W.J. and MAHER, J.V. 1967. Rainfall deciles as drought indicators. *Bureau of Meteorology Bulletin*, No. 48. Melbourne, Commonwealth of Australia.

GLANTZ, M. (ed.). 1994. *Usable science: food security, early warning, and El Niño*. Proceedings of the Workshop on ENSO/FEWS, Budapest, Hungary, October 1993. Nairobi and Boulder, Colo., UNEP and NCAR.

GLANTZ, M.; KATZ, R. and NICHOLLS, N. (eds.). 1991. *Teleconnections linking worldwide climate anomalies*. Cambridge, Cambridge University Press.

HEDDINGHAUS, T.R. and SABOL, P. 1991. A review of the Palmer drought severity index and where do we go from here? In: *Proceedings of the Seventh Conference on Applied Climatology*, 10–13 September 1991, Salt Lake City, Utah. Boston, Mass., American Meteorological Society, pp. 242–6.

HOKE, J.P. Jr. 1989. Savannah River Basin drought contingency planning. In: *Proceedings of the ASCE National Water Conference*, pp. 188–95. Newark, Del., 17–20 July 1989. New York, ASCE.

HREZO, M.S.; BRIDGEMAN, P.G. and WALKER, W.R. 1986. Managing droughts through triggering mechanisms, *Journal AWWA*, Vol. 78, No. 6, pp. 46–51.

IPCC (Intergovernmental Panel on Climate Change). 2001. Climate change 2001: the scientific basis. In: J.T. Houghton, Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell and C.A. Johnson (eds.). *Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, Cambridge University Press.

JOHNSON, W.K. and KOHNE, R.W. 1993. Susceptibility of reservoirs to drought using Palmer index. *Journal of Water Resources Planning and Management*, Vol. 119, No. 3, pp. 367–87.

KARL, T.R. and KNIGHT, R.W. 1985. *Atlas of monthly Palmer hydrological drought indices (1931–1983) for the contiguous United States*. Asheville, N.C., National Climatic Data Center (Historical Climatology Series 3–7).

KOGAN, F.N. 1995. Droughts of the late 1980s in the United States as derived from NOAA polar-orbiting satellite data. *Bulletin of the American Meteorological Society*, Vol. 76, No. 5, pp. 655–68.

MATHER, J.R. 1985. Drought indices for water managers. *Publications in Climatology*, Vol. 38, No. 1, pp. 1–77.

MAUNDER, W.J. 1992. *Dictionary of global climate change*. New York, Chapman and Hall.

MCKEE, T.B. and DOESKEN, N.J.; KLEIST, J. 1993. The relationship of drought frequency and duration to time scales. In: T.B. McKee, N.J. Doesken and J. Kleist (eds.). *Preprints, 8th Conference on Applied Climatology*, pp. 179–84. 17–22 January 1993, Anaheim, Calif. Boston, Mass., American Meterological Society.

NAULT, R.J.; GAEWSKI, P. and WALL, D.J. 1990. *Drought indication and response: optimizing the resources for water management*. In: R.M. Khanbilvardi and T.C. Gooch (eds.). *Proceedings of the 17th Annual National Conference*, 17–21 April 1990, Fort Worth, Texas. New York, ASCE, pp. 720–25.

NOAA. 1994. *El Niño and climate prediction: reports to the nation on our changing planet*. A publication of the University Corporation for Atmospheric Research pursuant to National Oceanic and Atmospheric Administration Award No. NA27GP0232–01. Boulder, Colorado, UCAR.

OGALLO, L.A. 1994. Validity of the ENSO-related impacts in Eastern and Southern Africa. In M. Glantz (ed.). *Usable science: food security, early warning, and El Niño*, Proceedings of the Workshop on ENSO/FEWS, Budapest, Hungary, October 1993. Nairobi and Boulder, Colo., UNEP and NCAR, pp. 179–84.

OTA (Office of Technology Assessment). 1993. *Preparing for an uncertain climate*, Vol. I. OTA-O-567. Washington, D.C., US Government Printing Office.

PALMER, R.N.; KUTZING, S.L. and STEINEMANN, A.C. (eds.). 2002. *Developing drought triggers and drought responses: an application in Georgia*. Proceedings of the EWRI Symposium in Water Resources Planning and Management, 19–22 May, Roanoke, Va.

PALMER, W.C. 1965. *Meteorological drought*. Washington, D.C., US Department of Commerce Weather Bureau (Research Paper No. 45).

RANEY, D.C. 1991. A water availability index for reservoir management. In: P.A. Krenkel (ed.), *Proceedings of the Environmental Engineering Conference*, 10–12 June 1991, Reno, Nev. New York, ASCE, pp. 686–91.

SHAFER, B.A. and DEZMAN, L.E. 1982. Development of a surface water supply index (SWSI) to assess the severity of drought conditions in snowpack runoff areas. In: *Proceedings of the 1982 Western Snow Conference*, 1982, Reno, Nev. Fort Collins, Colo., Colorado State University, pp. 164–75.

SMITH, D.I. HUTCHINSON, M.F. and MCARTHUR, R.J. 1993. Australian climatic and agricultural drought: payments and policy. *Drought Network News*, Vol. 5, No. 3, pp. 11–2.

TITLOW, J.K. III. 1987. A precipitation-based drought index for the Delaware River Basin, *Publications in Climatology*, Vol. 40, No. 2, p. 68.

USACE (US Army Corps of Engineers). 1991. *The national study of water management during drought: a research assessment*. Alexandria, Va., Institute for Water Resources (IWR Report 91-NDS-3).

WERICK, W.J. and WHIPPLE, W. Jr. 1994. *National study of water management during drought: managing water for drought*. Alexandria, Va., USACE, Water Resources

Support Center, Institute for Water Resources (IWR Report 94-NDS-8).

WHITE, D.H. and O'MEAGHER, B. 1995. Coping with exceptional droughts in Australia. *Drought Network News*, Vol. 7, No. 2, pp. 13–17.

WILHITE, D.A. 1995. Developing a precipitation-based index to assess climatic conditions across Nebraska. Lincoln, Nebr., final report submitted to the Natural Resources Commission.

WILLEKE, G.; HOSKING, J.R.M.; WALLIS, J.R. and GUTTMAN, N.B. 1994. *The national drought atlas*. Alexandria, Va., USACE (Institute for Water Resources Report 94-NDS-4).

WRIGHT, J.R.; HOUCK, M.H.; DIAMOND, J.T. and RANDALL, D. 1986. Drought contingency planning. *Civil Engineering Systems*. Vol. 3, No. 4, pp. 210–21.

## Additional References (Further Reading)

BLAIKIE, P.; CANNON, T.; DAVIS, I. and WISNER, B. 1994. *At risk: natural hazards, people's vulnerability, and disasters*. London, Routledge.

DAVIS, C.P.; WATSON, R.M. III and HOLLER, A.G. 1989. Southeastern drought of 1988: a federal perspective. In: T.A. Austin (ed.). *Proceedings of the National Water Conference*, 17–20 July 1989, Newark, Del. New York, ASCE, pp. 172–9.

FISHER, S. and PALMER, R.N. 1997. Managing water supplies during drought: triggers for operational responses. *Water Resources Update*, Vol. 3, No. 108, pp. 14–31.

HAYES, M. 1998. *Drought indices*. Lincoln, Nebr., National Drought Mitigation Center. www.drought.unl.edu/whatis/indices.htm (accessed 15 November 2004).

KNUTSON, C.; HAYES, M. and PHILLIPS, T. 1998. *How to reduce drought risk: a guide prepared by the Preparedness and Mitigation Working Group of the Western Drought Coordination Council*. Lincoln, Nebr., National Drought Mitigation Center. www.drought.unl.edu/plan/handbook/risk.pdf (accessed 15 November 2004).

PALMER, W.C. 1968. Keeping track of crop moisture conditions, nationwide: The new crop moisture index. *Weatherwise*, No. 21, pp. 156–61.

PHILANDER, S.G. 1990. *El Niño, La Niña, and the Southern Oscillation*. San Diego, Calif., Academic Press.

WILHITE, D.A. and VANYARKHO, O. 2000. Drought: pervasive impacts of a creeping phenomenon. In D.A. Wilhite (ed.). *Drought: a global assessment*, Vol. I, Ch. 18. London, Routledge.

## Appendix D: Flood Management


1. Introduction 603

2. Managing Floods in the Netherlands 605
   - 2.1. Flood Frequency and Protection 605
   - 2.2. The Rhine River Basin 605
   - 2.3. Problems and Solutions 609
   - 2.4. Managing Risk 609
     - 2.4.1. Storage 610
     - 2.4.2. Discharge-Increasing Measures 612
     - 2.4.3. Green Rivers 614
     - 2.4.4. Use of Existing Water Courses 615
     - 2.4.5. The Overall Picture 615
   - 2.5. Dealing With Uncertainties 615
   - 2.6. Summary 617

3. Flood Management on the Mississippi 617
   - 3.1. General History 619
   - 3.2. Other Considerations 623
   - 3.3. Interactions Among User Groups 624
   - 3.4. Creating a Flood Management Strategy 626
   - 3.5. The Role of the Government and NGOs 626

4. Flood Risk Reduction 627
   - 4.1. Reservoir Flood Storage Capacity 627
   - 4.2. Channel Capacity 630
   - 4.3. Estimating Risk of Levee Failures 631
   - 4.4. Annual Expected Damage From Levee Failure 633
     - 4.4.1. Risk-Based Analyses 634

5. Decision Support and Prediction 635
   - 5.1. Floodplain Modelling 636
   - 5.2. Integrated 1D–2D Modelling 637

6. Conclusions 638

7. References 640

# Appendix D: Flood Management

**Floods are always in conflict with floodplain development. No amount of structural protection will remove the risk of being damaged by flood flows and the mud, debris and pollution that accompany them. Hence the challenge is to be prepared to manage floods and mitigate the resulting damage when floods occur. In short, this means allowing space for floods when they occur and keeping that space available when they are not occurring. The challenge is to find the mix of open space and development on floodplains that maximizes the net expected monetary, environmental and social benefits derived from them. This chapter discusses some floodplain management approaches being considered in Europe and in North America, both motivated by the extreme flood events that occurred in the mid-1990s and early 2000s. Time will tell if any of these approaches are successfully implemented.**

## 1. Introduction

Just as droughts are defined as unusual scarcities or deficits of water, floods are defined as unusual surpluses or excesses of water. The result is higher than usual water levels that extend out onto floodplains. They occur in river basins, usually as the result of a lot of rain. They occur along coastal areas, usually as the result of severe storms and occasionally as the result of earthquakes. Floods are natural events. Flood damage usually results because humans and their structures occupy the space needed to store the excessive amounts of water. When a flood comes, as it will, these objects get inundated. To the extent that structures occupy some of the space floods need, flood levels increase, thereby flooding larger areas. The occurrence of flood events is not often enough to keep reminding people that floodplains can and do get wet. Flooded buildings and sandbags, such as shown in Figure D.1, are symbolic of what happens on occasion when floodplains are used, and often beneficially used, for other activities incompatible with flood prevention.

In mid-August 2002 Central Europe received an abnormal amount of rain. Floodwaters inundated floodplains along three major rivers, claiming around 100 lives and causing over 10 billion euros of damage to property in Germany, Russia, Austria and the Czech Republic. The upper left image in Figure D.1 illustrates a small but particularly damaging part of that flood in the historic portion of Dresden, Germany. The map in Figure D.2 shows the affected rivers.

The top image in Figure D.1 shows the River Elbe (Figure D.2) flooding the historic portion of Dresden in mid-August 2002. This flood of record exceeded the highest of floods in 157 years of records. Central Europe was not the only place getting flooded in the summer of 2002. By August of that year, flooding across Asia had claimed an estimated 1,800 lives. The waters of Dongting Lake and the Xiangjiang River (Figure D.3), which flows through the provincial capital of Changsha, were near all-time highs. If the sandbag dykes (middle image of Figure D.1) had failed around the 2,800 square-kilometre (1,070 square-mile) Dongting Lake that acts as a giant overflow for the flood-prone Yangtze River, over 10 million people and 667,000 hectares (1.6 million acres) of fertile crop land would have been flooded.

In the summer of 2002, millions of people in Bangladesh, India, Nepal, Thailand and Vietnam were

**Figure D.1.** Scenes of flooding, the efforts made to prevent flood damage, and coping at times of high flows and water levels.


**Figure D.2.** The Rivers Danube and Elbe in Central Europe experienced their most extreme flood of record in August 2002.


**Figure D.3.** Map showing Dongting Lake, which receives excess flow from the Yangtze River in China.

displaced from their homes by flooding. At least 900 people died in eastern India, Nepal and Bangladesh just in July and August of 2002 after heavy monsoon rains triggered widespread flooding, landslides and disease. In Venezuela at least 45,000 people lost their homes in flooding in the southwestern state of Apure as rivers overflowed and created lakes of stagnant, polluted water. In Algiers, the capital of Algeria, the flood toll reached 597 lives when muddy waters swept down a main road. In India's state of Assam, thousands of homes were destroyed by floodwaters. In Cambodia, water levels on the Mekong River rose to above emergency levels in two central towns following heavy rains upstream. Floods from rain falling across southern China inundated Vietnam's northern provinces, including streets in the capital Hanoi (lower image in Figure D.1).

In spite of considerable study and advice over the past half century on how to plan and manage floods, and in spite of increasing amounts of money being spent on flood protection, annual flood damage is increasing almost everywhere. In some years, like 2002, the damage can be quite substantial.

Of the major floods that occurred in the last decade of the last century, one (in 1993) occurred on the Mississippi River and some of its tributaries in the central United States, and two (in 1993 and 1995) occurred on the Rhine River. Both rivers are major multipurpose waterways serving many different interests, including barge transport, which is critical to the economy of the regions they flow through.

How can floods be managed? How can humans live with floods that, some believe, are becoming more frequent and more severe? This appendix discusses approaches to floodplain management in the Netherlands and in some parts of the United States. In both cases the goal is to avoid the continuing spiral of increasing costs, both for flood protection and then for flood damage recovery when the protection measures fail. (Additional information on flood modelling and risk assessment is in Chapters 7 and 11.)

## 2. Managing Floods in the Netherlands

Much of the land surface in the Netherlands is below sea level; thus, it is not a surprise that those who live in the Netherlands place a high priority on safety against flooding.

### 2.1. Flood Frequency and Protection

Levees (dykes) protect the economically important low-lying part of the Netherlands: roughly the western half of the country. The design levels of these levees are linked to the frequency of occurrence of a certain flood stage. The particular frequencies of occurrence, or risk levels, are determined by the Dutch Parliament. Translating those risks to levee heights is the job of engineers.

Levees along the coasts of densely populated and highly industrialized parts of the country are to be designed to protect against all storms whose magnitudes would be exceeded only once in 10,000 years on average. In other words, the levees could fail but only for storms that exceed that 10,000-year storm. The probability of such a failure happening in any year is 1/10,000 or 0.0001. Of course, there is no guarantee it could not happen two successive years in a row. On the basis of methods discussed in Chapter 7, the probability of having at least one storm that exceeds the design capacity of the levee at least once in a fifty-year period is 0.005. While considerably greater than the probability (0.0001) of it happening in any particular year, this is still a low probability. The Dutch are not risk prone when it pertains to floods.

For the less densely populated coastal areas, the design risk level is increased to storms expected once in 4,000 years, that is, those having a probability of 0.00025 of being exceeded in any given year. Along the Rhine and Meuse Rivers, the flood frequency is once per 1,250 years, or a probability of 0.0008 of being exceeded in any given year. These so-called design floods also constrain all landscape planning projects in the floodplain. Proposed river works for nature restoration, sand mining or any other purpose, need formal approval as laid down in the River Act.

The condition of flood control works, levees and fairways is monitored regularly. Every five years a formal report on flood safety is made. This involves re-determining the design floods using statistical analysis of river flows in the period 1900 to date. Furthermore, data regarding river cross sections and vegetation types and densities are updated. With the aid of that information, the design flood levels are assessed, taking into account effects of wind set-up and a freeboard margin of half a metre (twenty inches) for overtopping of the levee crests. If pre-established flood risk level tolerances are being exceeded, action must be taken to reduce these excessive risks.

While the principal objective is to protect places where people live, work and spend their leisure time, a secondary objective is to preserve the quality of the spatial environment, including natural as well as cultural and historical sites. The socio-economic interests of many sectors of society are considered when designing alternatives for flood risk reduction.

### 2.2. The Rhine River Basin

The Rhine, one of the largest rivers in Europe, extends 1,320 km through Switzerland, France, Germany and the Netherlands. It enters the Netherlands at Lobith and travels another 170 km until it reaches the North Sea. The Rhine catchment area, shown in Figure D.4, covers some 185,000 $km^2$, draining parts of nine countries.

The River Rhine now has a combined rainfall–snowmelt driven flow regime. The winter season shows the largest discharges originating from precipitation in the German and French parts of the basin. Summer discharges originate mainly from melting snow in the Swiss Alps when evaporation surpasses precipitation in the lowland region. Modern climate scenarios show a temperature rise combined with a rainfall increase during winter

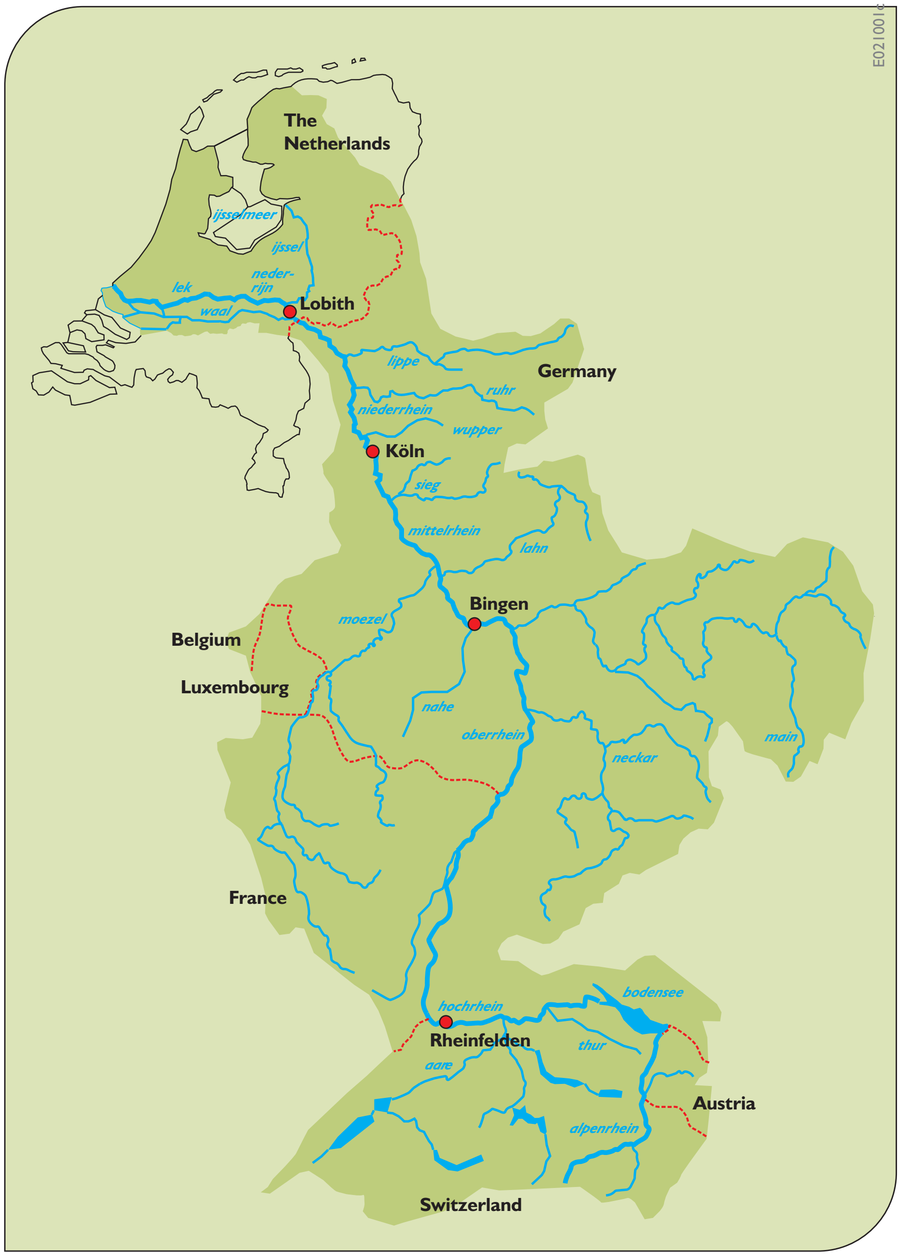



**Figure D.4.** River Rhine drainage basin. The river enters the Netherlands at Lobith.

in the basin. According to these scenarios the River Rhine could change from a combined snowmelt–rain-fed river into an almost completely rain-fed river. If that happens, both floods and dry spells could become more frequent.

To maintain the same flood safety standards, additional flood protection measures will be necessary. Water supplies in the future will be limited when they are most needed, and dry spells are likely to occur more often. Also, water quality may decrease due to both higher concentrations of pollutants and longer residence times of the water behind weirs. This means that preservation of water for use in dry periods could become necessary and that floodplains may be more frequently inundated.

The Netherlands is at the end of the Rhine and contains the Rhine Delta, which comprises more than half of the Netherlands (Figure D.4). A river often divides into multiple branches in its delta and the Rhine is no exception. In this case, the so-called Rhine Branches are the Waal, the Neder-Rijn/Lek and the Ijssel, as shown in Figure D.5. Along the Rhine branches, flood levels are entirely


**Figure D.5.** The Rhine branches and their floodplains in the Netherlands. The purple areas are subject to flooding and hence need dyke protection.

determined by the Rhine discharge. The area where the water levels are no longer determined solely by river discharge, but are also under the influence of the tidal effect of the sea, is referred to as the lower river region. In total, the low flow channel and the floodplains in the Netherlands cover an area of approximately 500 km$^2$.

To make and keep inhabitable that portion of the Netherlands that is below sea level, dykes have been built. As early as the mid-fourteenth century, an almost completely connected system of dykes arose that created the landscape of the Netherlands as it is today. There are two dykes on each side of a floodplain: the summer and the winter (flood season) dykes. These are shown in Figure D.6.

Design flow water levels in the river are mostly determined by the room that is present in the riverbed and by the flow resistance that the water encounters from

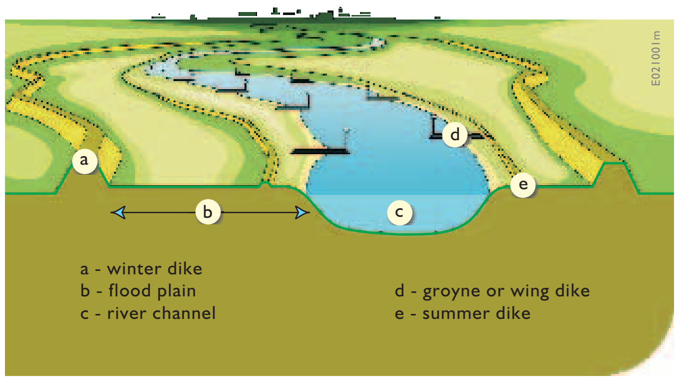



**Figure D.6.** Two-tier dyke structures and groynes for flood protection along the rivers in the Netherlands.

embankments and vegetation and other obstacles in the floodplain. The relation between design discharge and design water levels is therefore not entirely fixed, but depends upon the width and depth of the riverbed, the height of the floodplains, the flow resistance, and the wind and wave effects. Finally, an extra height is added to the dyke to equip it with an inspection path that keeps it passable, and to compensate for any subsidence of the embankment. In the lower river region, the design water levels are calculated with the aid of computer simulation models based on a large number of combinations of Rhine and Maas discharges, the sea level and the possible failure of the coastal flood barriers in the Rotterdam area.

The design height of a dyke is the water level whose chance of occurrence is linked to the level of protection that has been chosen for the protected regions. As already mentioned, practically all of the dykes along the Rhine Branches have a protection level of 1/1,250 per year. This means that the chance that the water level in a particular year will be higher than the design water level is less than 1/1,250. In the western part of the Netherlands, the protection levels are higher, namely 1/2,000, and rise to 1/10,000 per year for 'Central Holland,' where large cities such as Rotterdam and The Hague are located. Here, the population density and economic interests are larger, but so also is the difficulty of predicting a storm at sea. Floods from seawater can result in substantial casualties and economic damage due not only to water, but also to salt.

Two movable gates protect the entrance to the Rotterdam harbour, as shown in Figure D.7. Each gate is about as long as the Eiffel Tower is tall and contains over twice the tower's weight of steel.

In 1995 – after a long period of relative freedom from floods – the River Rhine region witnessed a flood. It was not only the highest since 1926, but also one that was long in duration. Approximately 250,000 people were evacuated for a little under a week due to the questionable stability of saturated dykes that had been exposed to long-term flooding. This flood – together with the flood of 1993 and the comparable events along the River Maas (Meuse), the second largest river in the Netherlands – motivated government agencies to give some priority to flood safety and flood risk reduction in the river regions.

Thanks in part to its location in the Rhine Delta, the Netherlands has been able to flourish economically. Rotterdam is one of the largest harbours in the world with a large and rich hinterland. Agriculture in the Netherlands has been able to profit from the rich soils that the Rhine has deposited. The economically strong position of the Netherlands is also due in part to the dykes. At the same time, the dykes have come to represent a sort of Achilles' heel for the country.

A large part of the Netherlands lies below the high-water level of the major rivers. At present, the sea level is gradually rising as a result of climate change. This same climate change, together with watershed land use changes within and upstream of the Netherlands, may cause the peak discharges on the Rhine and Maas to increase. In the meantime, the area protected by the dykes has been sinking, primarily due to subsidence and oxidation of peat, because the soil is so 'well' drained. This has increased the difference between water levels in the area protected by the dykes and the area outside of the dykes.

The population in the area protected by the dykes, the land use intensity and the capital investment have all rapidly increased. As a result, the potential adverse economic and emotional consequences of a flood, or even an evacuation due to flood risk, have increased substantially.

**Figure D.7.** A drawing and photograph of the waterway storm surge barrier in the closed position at Rotterdam in the Netherlands.

Yet there is a limit to just how much more heightening and reinforcing of dykes is reasonable. The public has made it clear they do not want dyke heights increased. Many feel too boxed in as it is, in this crowded country. Today a more natural solution is sought.

## 2.3. Problems and Solutions

As a result of the floods of 1993 and 1995, the design once-in-1,250-year discharge that must be contained or controlled within the floodplain is now higher than it was prior to these events. As shown in Figure D.8, the design discharge of the Rhine is now established at 16,000 $m^3$/s, an increase of 1,000 $m^3$/s. This design discharge height determines the design height of the dykes.

Without further measures, this means higher dykes. Due to climate change, the design discharge will most probably further increase to 18,000 $m^3$/s in the second half of this century. At the same time, the sea level is expected to rise, causing backwater effects in the estuaries and rivers.

Given this increase in the design water level, one could argue that the dykes should be made higher and stronger. If this is not done, the probability of floods topping the dykes will increase. However, the higher the dykes, the more severe the impacts will be once a dyke fails. The dykes often protect low-lying 'polders,' whose potential flooding depths already exceed several metres. If nothing is done to change the available room that exists for the Rhine River flood flow, this will lead to 20–30-cm higher design water levels in the short run, and up to 90-cm higher design water levels in the long run, factoring in climate change impacts. This will make higher dykes necessary in order to maintain the current level of flood protection. The higher the river water level is, the larger the consequences are, should a major flood occur that exceeds the dyke capacity. In reality, however, higher flood protection dykes are not going to be a problem, since the public has made it clear they are against that option. So, how to provide the extra protection needed to meet the design risk criteria established by the Dutch Parliament?

The challenge is to create and execute measures that – despite the higher design discharge that must be contained – avoid the need for additional dyke heightening. The only way to do that is to create more space for water to flow at existing dyke levels. This means increasing the widths or depths of the rivers or their floodplains, and providing flood overflow areas outside the dyked floodplain for use when the space within the dykes is full. Specific floodplain planning and designs needed to cope with the design flow of 16,000 $m^3$/s are to be implemented by 2015.

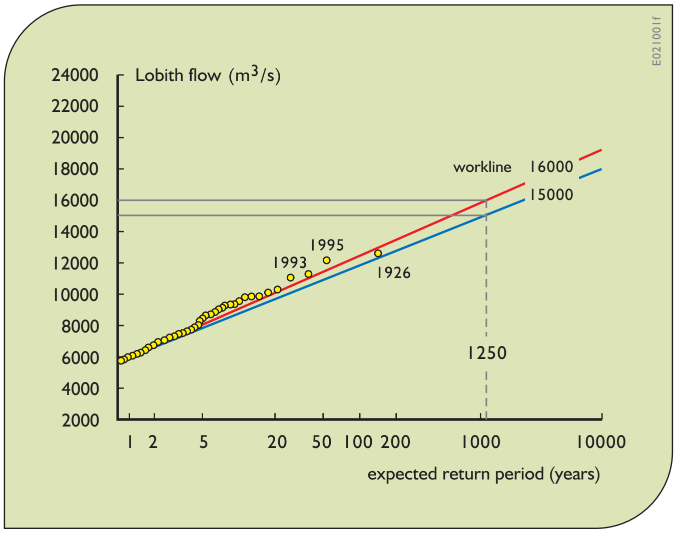



**Figure D.8.** Re-evaluation of design flood flows for the River Rhine in the Netherlands.

## 2.4. Managing Risk

There is always a possibility that the water level along a river will exceed the design dyke capacity. The desired level of safety is a matter of societal choice. If society wanted complete and total safety, the only way to achieve it would be to move out of the floodplain. For natural phenomena such as wind, rain and river discharge, there are no absolute known upper limits. There are, thus, no absolute protection guarantees; a finite risk of flooding will always exist. What society can do however is protect to the extent people deem appropriate, while preparing for flood events that exceed the design capacity of any implemented protection works. Controlled flooding in emergency overflow areas is one way to manage flood events that exceed the design water levels.

The notion of flood risk considers both the chance of a flood and its adverse consequences. Minimizing the chance of inundation is not the only possible risk reduction option. Diminishing the adverse consequences of the event also reduces risk. This idea in particular underlies the design of emergency overflow areas: it is better to have a controlled flood with minor damage than an uncontrolled flood with major damage.

It is technically possible to continue raising dyke heights, now and on into the future, just as has been done in the past. But Dutch society has made it clear that this is undesirable. Dyke reinforcements bring with them increasingly negative consequences for landscape, nature, and cultural and historical values. And as the difference in height between the water level in the river and the area protected by (behind) the dykes continues to increase, the flood risk also increases. The larger population and capital investment in the area protected by the dykes already make uncontrolled floods more drastic events than they were some twenty to fifty years ago. Moreover, the perception of safety increases as the dykes are made higher and more heavily reinforced, which leads to an increase in infrastructure investment, which leads to greater land value and hence potential loss should a flood occur, which justifies further flood protection investments, and so on.

It is for this reason that a policy is needed to break this cycle. Such a policy needs to:

- anticipate floods instead of merely reacting to them
- make more room for water, rather than relying only on technical measures such as dyke heightening
- prevent the transfer of water problems to downstream areas by means of a detain–store–discharge sequence.

This requires river-widening and deepening measures to prevent new design flood flow water levels from exceeding current dyke heights.

Past human interventions have resulted in the erosion of the low flow channel in upstream sections due to maintenance work on the navigation channel; the sedimentation in the low flow channel in the downstream sections after estuaries were closed off; and the silting-up of floodplains due to the construction of dykes and summer embankments. These processes have increased the elevation of the floodplains in relation to the low flow channel. At the same time, the area protected by the dykes has been subsiding due to drying out and oxidation of the peat soils. All of this has resulted in river levels that are higher above the surface of the land protected by (behind) the dykes than they are above the land adjacent to the river in front of the dykes.

There are several ways to maintain design water levels as design discharges increase in the Rhine:

- keep water in the catchment area upstream of the Netherlands
- store (excess) water along the Rhine Branches in the Netherlands
- discharge (excess) water via the Rhine Branches.

The first alternative attempts to ensure that the upstream precipitation does not lead to higher discharges downstream at a later time. This requires measures in the catchment area upstream of the Netherlands, namely Germany, such as changes in land use or detention. Comparable measures in the Netherlands may also help to reduce the additional flow to the Rhine Branches from tributary streams and canals.

Alternatives for storing and discharging floodwaters can be implemented in the Netherlands. The storage of river water in detention areas along the Rhine Branches leads to lower peak flows, lowering the water levels downstream of the detention areas. Conversely, measures that increase the discharge capacity of the riverbed reduce the water levels upstream of the measure. Examples include the removal of obstacles in the winter bed such as high lying areas, ferry ramps or bridge abutments, excavation of the floodplains, lowering of groynes or wing dykes, dredging of the low flow channel, and setting back the dykes. Some of these options are illustrated in Figures D.9, D.10 and D.11.

An important condition for such projects is that they must allow at least an increase in the design flow from 15,000 to 16,000 $m^3/s$. A further condition is that they cannot change the proportions of river flow that enter each of the Rhine Branches in the Netherlands.

### 2.4.1. Storage

Storage reduces discharge, and is helpful for practically the entire area downstream of the storage basin. Emergency overflow and temporary storage areas are being considered for controlled flooding in a way that limits damage. A segment of the discharge peak could be attenuated by temporarily storing it in a dyked area. After the flood peak has passed through, the temporarily stored water can be released.

Since the desired effect of a detention area occurs downstream from the area itself, a location as far upstream as

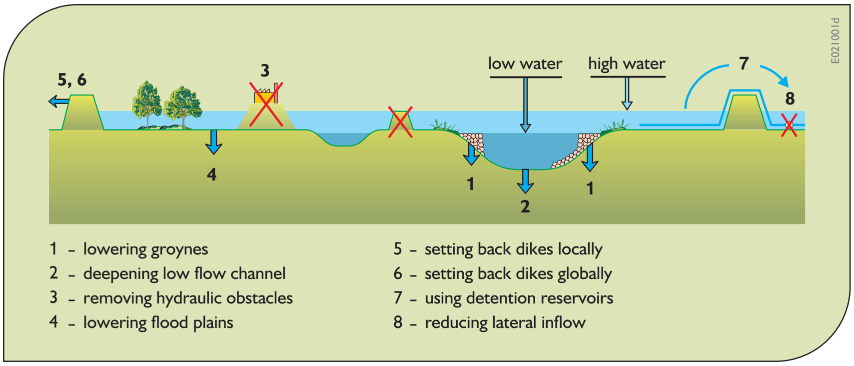



**Figure D.9.** Alternatives to dyke heightening for increasing flood flow capacity of rivers.

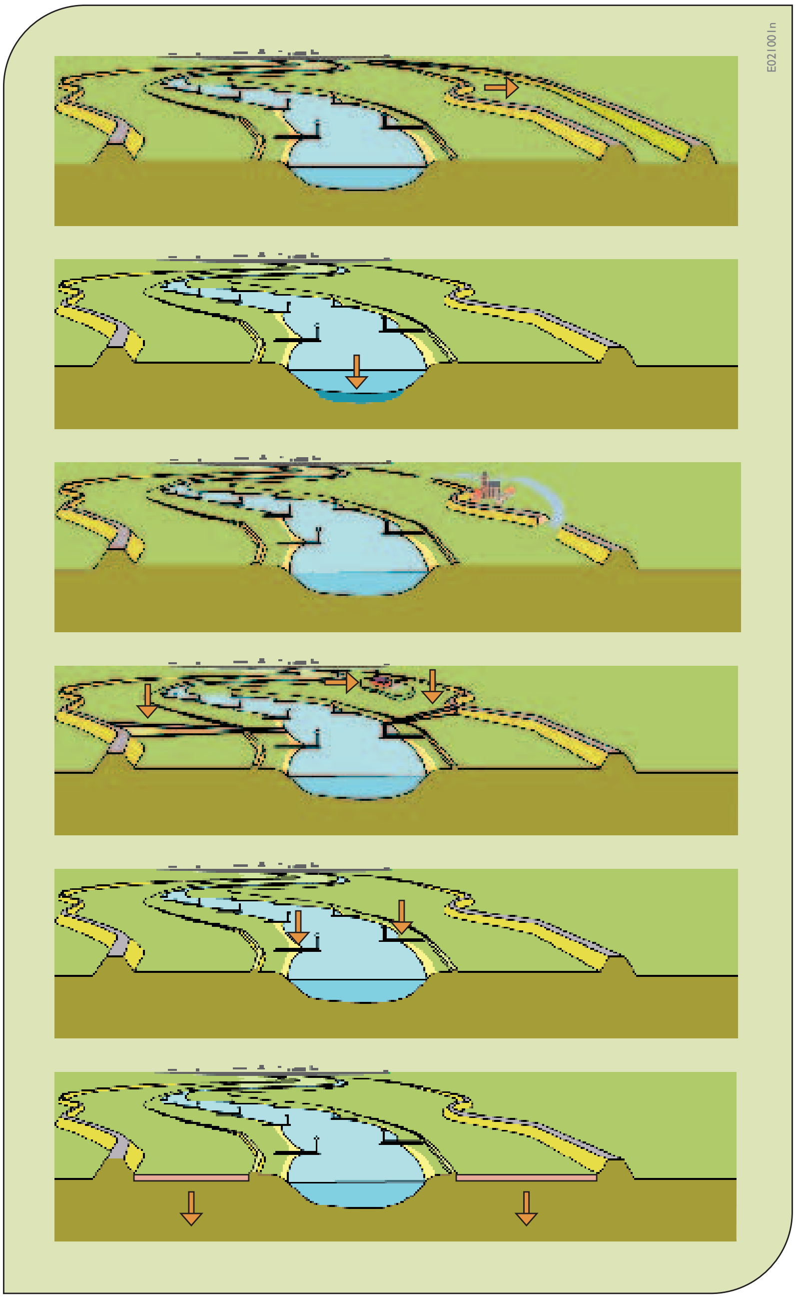



**Figure D.10.** Alternatives to dyke heightening are, from top to bottom, setting back dykes, deepening low flow channel, development of green rivers, removal of hydraulic bottle necks, lowering groynes, and lowering floodplains.

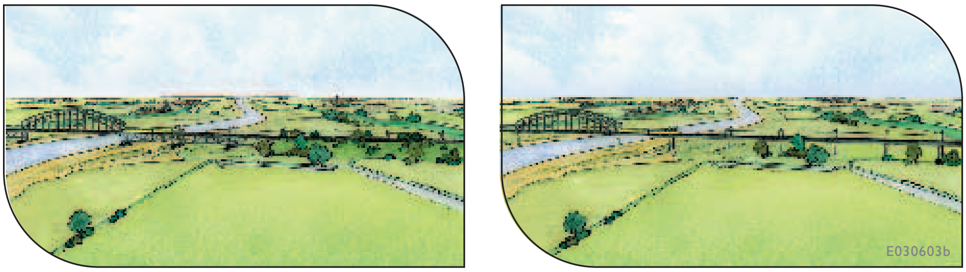



**Figure D.11.** Bridge abutments and supporting structure, before (left) and after (right) floodway modification.

possible is preferred. In the Netherlands, detention areas close to the German border are thus the most appealing.

The total detention (storage) capacity necessary to maintain a given flood height depends on the shape and duration (or length) of a flood wave. For example, a storage capacity of more than 150 million m$^3$ would be required to reduce the peak flow by 1,000 m$^3$/s of a flood wave lasting several days and having an 'average' shape. This in turn would require a detention area of some 3,000 hectares (30 km$^2$) flooded to a depth of five metres. In areas with less depth, the surface area must be proportionately larger.

Detention areas require enclosure dykes and intakes, both of which can be extensive. Implementing detention areas usually also involves landowner evacuation, and compensation for any damage occurring during occasional flooding. All of these factors add to the cost of detention measures.

Detention areas seldom need to be used, and in this case the probability of them being needed in any year is approximately 1/500, as seen from Figure D.8. This low probability can be a problem. The less often an area overflows, the more the societal pressure will be to develop it and safeguard it against flooding. After many years without floods, people may begin to think that these detention areas will never be needed for flood detention. The result: flood flow restricting activities will begin to encroach on that land. The additional flood protection provided by the detention area would then be decreased if not entirely lost. Perhaps this suggests the need to periodically let some water flow into these areas during high river flows, even when it might not have been necessary, and/or alternatively to allow it to become an open recreation area or nature reserve.

For detention areas to be effective during flood events, a reasonable amount of precision in their operation is required. They must not fill up too soon, because they run the risk of being full before the actual peak discharge has arrived. If they cannot store some of the peak flow, when it arrives, they will have no effect in lowering the maximum flood stage. A similar danger exists for a very lengthy, flattened flood wave, or when a second peak occurs soon after the first, and the detention area has not emptied.

Detention basins must also not fill up too late, because the discharge peak will already be over. All these considerations imply that accurate predictions for the timing and shape of the discharge must be available if detention basins are to be used effectively.

In the past, detention basins were usually filled via a fixed sill, a lowered section of the dyke that allows overflow to take place in a 'controlled' manner, ideally uniformly over the entire length of the sill. It is still very difficult to design a sill that will ensure a uniform overflow over a substantial length. This problem may be overcome by using a 'regulated' intake under human control. However, people have had bad experiences with regulated intakes. The residents of an area are never in favour of inundation, and if some person is making it occur a conflict with the responsible organization or agency, if not the operator, is inevitable.

### 2.4.2. Discharge-Increasing Measures

Measures taken to increase the discharge capacity can also reduce water levels while maintaining the same design discharge. In contrast to storage, however, increasing the discharge capacity is only advantageous for a limited river section.

With discharge-increasing measures, it is not only the height reduction but also the distance a measure covers that is important. The distance depends on factors such as the steepness of the water level slope, the location of the dykes on the floodplain, other obstacles in the area, the hydraulic roughness of the vegetation in floodplains and so on.

There are many measures available for reducing the stage of a particular design discharge. Three major ones include:

- increasing the flow capacity in the low flow channel
- increasing the flow capacity in the floodplains
- providing flood storage capacity in the areas protected by (behind) the dykes (by, for example, setting back dykes or creating detention basins).

As mentioned earlier, sedimentation occurs in downstream sections and this makes regular dredging necessary. Dredging to deepen the low flow channel in the downstream sections can lead to a water level reduction. However, it can also accelerate erosion upstream. Thus, to maintain the desired design water level, continual dredging may be required.

Alternatives to dredging include groynes or wing dykes (Figures D.6, D.9, and D.10). Groynes were constructed in the past to ensure that the river retained a sufficient depth without continual dredging. They guide the river flow to the middle of the channel and ensure that the depth of the river is maintained for a pre-determined width. This is particularly important for navigation. They also tend to prevent sand banks.

Removal of the groynes would result in a decrease in flow velocities and the river would become shallower in places. Sandbanks might even form in the middle of the river. With few exceptions shortening or removing groynes is an option only if the shipping function of the river were to be discontinued, and of course in the River Rhine it is not.

It appears from simulations that lowering of groyne heights could contribute to a reduction varying from 5 to 15 cm in the water level on the Waal and the Ijssel. On the Neder-Rijn, the maximum reduction is 10 cm. This may not seem like much, but on the other hand, the costs of groyne lowering are relatively low; thus, this measure is relatively cost effective.

Excavation of the floodplains and the removal of hydraulic bottlenecks can also be considered. Floodplain excavation is a measure by which the gradual height increase caused by sedimentation on floodplains may be counteracted. Floodplain excavation may be combined with clay mining, and dyke reinforcement, and/or with developing nature reserves. Nature development is often a reasonable use of an excavated floodplain since excavation makes land less profitable for agriculture, particularly if the summer embankments are also removed in the process. Moreover, it appears that after excavation, such development produces a result that is valued by many. Creating nature reserves without floodplain excavation, however, pushes water levels upward since rough vegetation, scrub, and wooded areas can slow down the discharge. Thus, it requires additional excavation to compensate for the backwater effect of the vegetation.

While floodplain excavation is an effective way to reduce flood heights, it is also the most expensive measure. If it were to be implemented along all three Rhine Branches, the costs involved would total 3 to 4€ billion. Soil excavation is expensive in and of itself, but it is the necessary storage and containment of contaminated soil that substantially increases the costs. Roughly 15 to 20% of the soil on the Rhine floodplains is actually contaminated and another 40 to 50% is unusable as building material.

Storing the contaminated soil safely and locally in existing deep ponds or in sand excavation pits after usable material has been removed, so-called 'earth-swapping,' can achieve substantial cost savings, from 1 to 1.5€ billion.

The removal of hydraulic obstacles in the floodplain is another way to increase its discharge capacity without increasing its water level. Examples of hydraulic bottlenecks include ferry ramps, bridge abutments (Figure D.11), high-lying areas (Figure D.12), summer

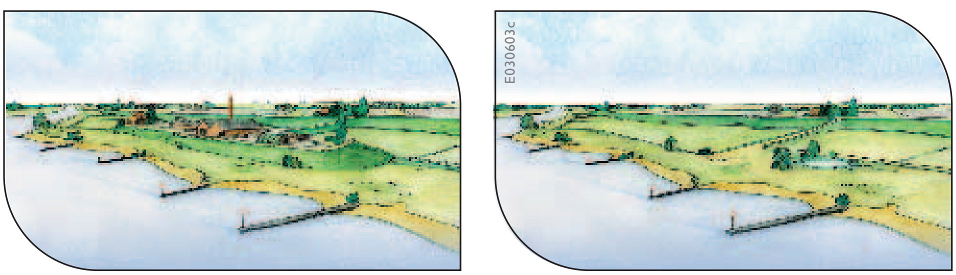



**Figure D.12.** Removal of high-lying areas on a floodplain.

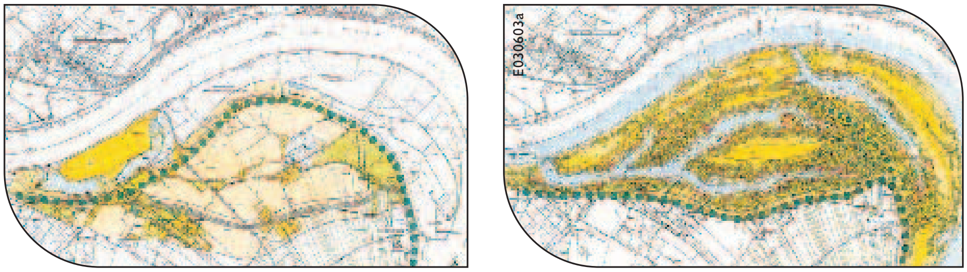



**Figure D.13.** Opening up a narrowing by setting dykes further back, and at the same time creating a nature reserve.

embankments that are high and/or perpendicular to the flow direction, narrowing of winterbeds (Figure D.13) and other obstacles. Hydraulic bottlenecks may be identified by studying the water level slope of the river. Typically, there is a direct relation between structures and a change in the water level slope.

Removing bottlenecks can decrease design flow water levels. This decrease can be calculated, as can its cost. The costs of replacing bridge abutments by bridge sections (Figure D.11) and the removal of ferry ramps vary from less than 2.5€ million to more than 75€ million for a large bridge. The costs of removing embankments and small-scale setting back of dykes (Figure D.13) are usually on the order of 5€ million per project, but they can run up to over 15€ million if many houses must be expropriated. The costs for removing high-lying areas (Figure D.12) can amount to 30€ million, especially in those cases where land is contaminated.

Substantial water level reductions may be achieved by widening and deepening measures at an urban bottleneck. Such measures are typically very expensive, as a sizeable area is often needed in times of flooding. Despite their high costs, these measures can be cost effective at the urban bottlenecks due to the relatively large reduction in water levels that may be achieved.

On average, the removal of about sixty bottlenecks can reduce the water level by 20 cm on the Waal and 10 cm on the Neder-Rijn/Lek and the Ijssel. However, actual water level reductions may vary considerably over the length of the river branches.

There are some forty locations, not including urban bottlenecks, where large-scale setting back of dykes can lead to a substantial decrease in the water levels. Setting back of dykes is particularly effective in situations where the winter bed is very narrow and causes backwater effects quite a distance upstream. In such a case, this decrease in the water level also continues to work relatively far upstream. Some setting back of dykes can, in fact, decrease the water level up to half a metre. There are other sites where such measures produce only several centimetres worth of reduction.

To set back a single dyke section of a few kilometres in length costs anything from 5€ million to 100€ million. While rather expensive, particularly if considerable urban or industrial development is present on or just behind the dyke, these measures are cost effective compared to many other hydraulic bottleneck reductions. Along the Waal and the Neder-Rijn/Lek, the setting back of all the dykes together could result in a maximum reduction of 60 cm.

### 2.4.3. Green Rivers

If it is not possible to set back a dyke, or if the effect is too small, then a so-called 'green river alternative' may be an effective way to reduce flood levels. Green rivers are floodplains between two dykes where water would flow only during floods. Green rivers may be used for agricultural purposes or may be designed for nature reserves and/or recreational areas; they are, in short, 'green.' This does not preclude the possibility of digging a channel or lake into such an area for the sake of recreation, for example. How such a green river may be designed depends upon the location. Figure D.14 shows a section of a river floodplain converted to a green river.

Green rivers can lead to significant reductions in water levels at and upstream of those locations. This means that other flood height reducing measures along these river sections may not be necessary.

Green rivers offer options for agriculture, water-based recreation and nature development. The land is seldom flooded and, if it is, this usually occurs 'off-season' (outside the agricultural season) just as it currently does in the floodplains. One could say that the land is temporarily loaned to the river every year, but is otherwise available

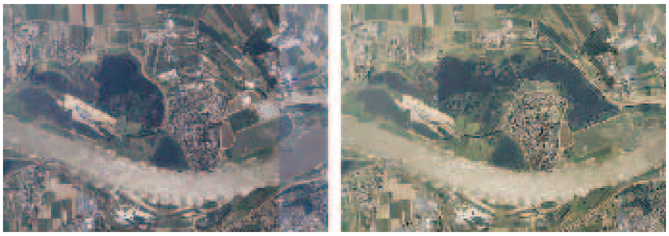


**Figure D.14.** Existing situation (left) of the Rhine floodplain near the Dutch town of Haaften, and artist's impression (right) of a green river north of the town.

for other compatible activities. There is thus a definite practical value derived from green rivers.

Additionally, these measures offer what one could refer to as a future value for river management. Efforts must be made to prevent such space from being used for incompatible residential construction, business parks, greenhouse complexes and similar land developments that would incur substantial damage, and indeed increase the flood height, should a flood occur. While this limits the land use possibilities at this moment, it offers the option to implement other river widening and deepening measures in the future, such as floodplain excavation.

Similar far-reaching measures may be taken to improve the quality of the surroundings, such as clean-ups of industrial areas and development of recreational areas.

### 2.4.4. Use of Existing Water Courses

Where existing canals, brooks or creek remnants run parallel to the river, these may be connected to the river channel to take on a portion of the discharge. Opportunities to do this have been examined in the lower Rhine Branches region in particular. New, still-to-be excavated channels are also being studied with respect to their effectiveness and costs. In practically all cases, however, water is guided in an entirely different direction and this places increasing burdens on other rivers or sections situated further downstream.

### 2.4.5. The Overall Picture

It would appear that the large-scale setting back of dykes, construction of detention basins/green rivers and lowering of groynes would result in the greatest water level reduction per euro invested. The removal of hydraulic bottlenecks obtains an average score, as does dredging the low flow channel. Floodplain excavation is the most expensive, and in this respect, the least desirable type of measure.

Some measures, such as lowering of groynes and floodplain excavation, are only really possible in upstream sections. Others, such as dredging of the low flow channel, are more feasible downstream. Large-scale setting back of dykes and green rivers relieve certain bottlenecks only, albeit with substantial carry-over upstream.

Finally, cost effectiveness is only one criterion. Sometimes floodplain excavation can involve multiple objectives; nature development and even sand and clay mining may also profit from it. The extent to which similar multiple objectives may be served by various flood capacity enhancing alternatives needs to be explored.

Combinations of flood-height-reducing measures will undoubtedly be undertaken along the Rhine River Branches. Models will be needed to assess their overall effectiveness. The overall reduction in flood heights will not be simply the sum of their individual reductions in isolation. It is not possible to simply add up the water level reducing effects of the different measures. The discharge of a river is, after all, determined by the functioning of the whole: one single bottleneck can negate the effect of a package of measures. On the other hand, some measures are synergistic; their overall effectiveness can be greater than the sum of all their individual water level reductions. For this reason, a systems view is necessary to effectively lower water levels across the entire length of the Rhine Branches. Not only the combination of measures but also the possible changes in the shape of the flood wave must be taken into account.

There are an innumerable number of alternatives that could be put together to safely contain the design flow of 16,000 $m^3$/s. It is a matter of preference which measures will be applied first or most often.

To handle a discharge of 18,000 $m^3$/s safely, it appears that large-scale measures in the dyke-protected area would be necessary, such as setting back dykes and creating detention areas and green rivers.

## 2.5. Dealing With Uncertainties

There are many uncertainties in predicting flood levels, and sometimes there is little one can do about them. One uncertainty involves the design discharge itself. Others arise over the shape of the flood wave, the distribution of River Rhine flow over the three Rhine Branches, the bed level of the river and the roughness of the vegetation in the floodplains along the three branches. All these uncertain factors affect the design water levels; hence, the design water level is itself uncertain.

Moreover, we are at the mercy of changes that may occur in future and these are by definition uncertain. We know the climate will change, but we do not know how quickly or to what extent. All the climate models currently used predict warming and more frequent extremes, but the variations between predictions remain rather substantial. All of these issues present quite a dilemma for management: on the one hand, safety is so vital that the river manager has to anticipate higher discharges that must be accommodated, yet the speed with which situations may change is very unpredictable. To wait and see how new floods influence design flow risk statistics seems unacceptable. Further research can do nothing to change that.

Essentially, decisions must be made in spite of these uncertainties.

The design discharge for the Rhine is based upon an extrapolation of measurement data from the past. Previously, the period was 1901 to 1991, but this has now been extended to 1995. The extension of only four years, two of which had high discharges (namely 1993 and 1995), has caused the slope of the graph to change somewhat, as shown in Figure D.8, above. The new design discharge has become 16,000 m$^3$/s, an increase of 1,000 m$^3$/s. It appears that several floods resulting from relatively wet years can substantially influence the design discharge.

The most important reason for the large change in the design discharge is that it applies for events that occur once every 1,250 years, well beyond what has been observed during the 100 years of measurements. This means that the graph must be extrapolated beyond the measured data to estimate the design flow associated with that 1/1,250 risk level. This can result in strange effects. For example, a plot of the annual peak flows of the Oder River catchment area in Poland from 1901 to 1985 produces a fairly straight line without any large deviations. In 1997 however, a discharge of 3,300 m$^3$/s was measured, the largest flow on record. That one event led to the 'new' extrapolation line causing the 1/1,250 discharge to rise from 2,500 to 2,600 m$^3$/sec. This is still far below the 3,300 m$^3$/s discharge that actually occurred. While extreme discharges, perhaps such as that on the Oder, may be rare, they can nonetheless occur in any year.

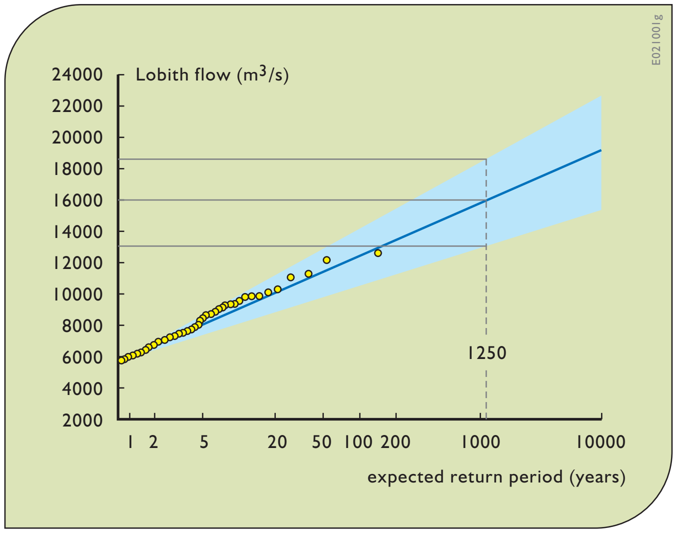



**Figure D.15.** Range of uncertainty associated with estimated design flows of various return periods. The true design flow has a 90% chance of being within the blue band.

What is clear is that any extrapolation line, such as the one shown in Figure D.8, is uncertain. It could be higher, or it could be lower. Figure D.15 shows the 90% confidence interval band of uncertainty associated with the extrapolation in Figure D.8. This shows that there is a 90% chance that the design discharge on the Rhine (that is, expected on average once in 1,250 years) will be between 13,000 and 18,500 m$^3$/s. Not only the new design flow of 16,000 m$^3$/s but also the old design flow of 15,000 m$^3$/s, as well as the possible future design flow of 18,000 m$^3$/s, are in this range.

Figure D.15 is based on historical flow data that may not be indicative of future flows. There are various climate change scenarios, each with high and low estimates. Assuming a high estimate of a 4 °C rise in temperature in the year 2100, the design discharge on the Rhine could increase by 20%. This added to the current 16,000 m$^3$/s would produce a design discharge of more than 19,000 m$^3$/s, assuming that Germany is successful in keeping this discharge within the dykes.

In this respect, there are thus two types of uncertainties: first, what precisely would change in terms of climate, and second, how the other Rhine States upstream of the Netherlands would react to these changes. No one today can answer that.

How can flood management proceed, given this uncertain future? The answer is: by building into any adopted management strategy both flexibility (robustness) and resilience.

Flexibility is the ability to adapt with minimal cost to a wide range of possible futures. Building in this flexibility may cost more, but may be still be desirable insurance against risks society does not want to take. Regulating the discharge distribution over the Rhine Branches could increase flexibility. So too could building temporary and emergency detention areas.

The term resilience, on the other hand, specifically involves the speed of recovery after a flood and its accompanying damage has occurred. This is achieved much more easily if the consequences of an above design level flood are not permanent, but remain limited and may be easily rectified. This requires that no uncontrolled flood should occur – one that would be accompanied by possibly extremely severe damage or even social disruption – but only controlled flooding that will cause the least amount of damage. In this manner, resilience could be 'built into' a flood safety system through disaster facilities, for example in the form of emergency spill areas or by dividing large dyke rings up into smaller sections – compartmentalizing – to limit flood damage.

It is a major challenge for river managers and also for the water and spatial-planning policy-makers to develop a strategy that will minimize future pain by taking into account the fact that there will always be uncertainties about the expected discharges in rivers.

### 2.6. Summary

Visions for future developments on the River Rhine in and upstream of the Netherlands currently concentrate on flood mitigation and ecological restoration. Relatively little effort is devoted to dealing with low flows that have implications for future water quality and navigation requirements.

Until recently, the strategy for flood prevention was to raise dykes (embankments) along the floodplains. Currently this strategy has met social resistance and is thought to be too inflexible to cope with an uncertain future. Alternative solutions focus on reducing water levels during floods, using retention basins along the River Rhine in Germany and lowering floodplains in the Netherlands to enlarge the cross section of the river. At the same time, these floodplains are to be designed in such a manner that more natural morphological and ecological processes in the floodplains can take place.

In the delta of the River Rhine, future adaptation visions focus on a further widening of the floodplains and the planning of 'green rivers.' These 'green rivers' will only be inundated during floods. In the upstream sections in Germany, the focus is on landscape planning so that water will flow less quickly to the river.

## 3. Flood Management on the Mississippi

The Mississippi River, shown in Figure D.16, is the major river of North America and the United States. It is over 3,700 km (2,300 miles) in length, flowing from the northwestern part of the state of Minnesota south to the Gulf of Mexico, just below the city of New Orleans. It drains 3.2 million km$^2$ (1.2 million square miles), or about 40% of the United States. The Mississippi is the sixth largest river in the world in terms of discharge. Its annual average flow rate is 14,000 m$^3$/s and its freshwater discharge onto the continental shelf is 580 km$^3$/yr. It is a significant transportation artery, and when combined with its major tributaries (the Missouri and Ohio rivers), it is the third largest river transportation system in the world.

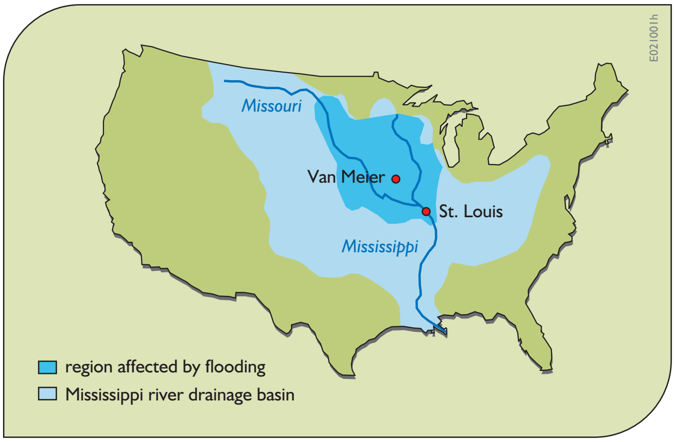



**Figure D.16.** Mississippi River Basin in the United States and the area affected by the 1993 flood.

**Figure D.17.** Confluence of Des Moines and Mississippi Rivers during the 1993 flood.

Typically the discharge at the mouth of the Mississippi River into the Gulf of Mexico varies over the seasons, with the highest flow occurring between March and May and the lowest between August and October. In August and September of 1993, however, it rained heavily. Monthly mean discharges in April and May 1993 were approximately 50% higher than their long-term monthly mean values. August and September 1993 discharges exceeded the highest of those recorded in the previous sixty-three years. The result was a flood.

The States of Minnesota, Iowa, Illinois, and Missouri were hardest hit (the darker blue area of Figure D.16). At St. Louis, the river crested at about 15 m (49.6 ft), some 6 m (over 19 ft) above flood stage. This was about two metres (more than six feet) above the old record set in 1973. The Mississippi remained over flood stage at St. Louis for over two months. Farther north, record flooding occurred on the Des Moines River, a tributary of the Mississippi (Figure D.17). At one point flooding disabled a major water plant, and Des Moines, Iowa, a city of nearly 200,000 people, was without safe drinking water.

The discharge of 35,700 $m^3/s$ (1,070,000 $ft^3/s$) was the greatest ever measured during more then 130 years of data, exceeding the previous high by 26%. Flood elevations exceeded the design flood stage of 9 m (30 ft) on the St. Louis gauge for eighty consecutive days during the main portion of the flood, and for 148 days during the calendar year. The previous record was seventy-seven days above flood stage, both consecutively and annually.

The duration of flooding at high stages was unprecedented. The flood was 3 m (10 ft) or more over the design flood stage for thirty-six days, exceeded the 'fifty-year flood' stage for twenty-three days and exceeded the '100-year flood' stage for eight days. Before 1993, there had only been twelve days total in the entire period of record, dating back to 1861, when floods exceeded the design flood stage by this extent or more. The total water volume passing St. Louis from June 26 to September 13 could cover the entire State of Missouri to a depth of 0.75 m (2.5 ft). The peak discharge at St. Louis has been estimated to have a 150–200 year average recurrence interval. At some upstream stations the flood exceeded the 500-year average recurrence interval.

Transportation and industry along the Mississippi River was disrupted for months. Damage to surface and river transportation in the region was the worst ever incurred in the United States.

Over 1,000 of the 1,300 levees designed to hold back floodwaters failed, though major cities along the rivers, like St Louis, were protected from flooding by massive floodwalls. Floods displaced over 70,000 people, and damaged or destroyed nearly 50,000 homes and over 31,000 $km^2$ (12,000 square miles) of productive farmland. Fifty-two people died. The estimated damage was between \$12 and 20 billion (1993 US dollars), depending on what is included in those estimates.

This flood provided a trigger, once again, for an evaluation of the existing flood protection schemes and procedures for managing the rivers and floodplains in the Mississippi Basin. At the same time, the future use of the rivers and floodplains, especially in the context of economic development, coupled with the improvement of ecological habitats, became important political issues.

Historically considerable attention has been given to solving problems associated with particular interests. For the Mississippi Basin, these include navigation, hydropower generation, nature conservation and providing a certain flood safety for agricultural, industrial and urban areas. A large number of structural and non-structural measures have been implemented over the years, many of which have contributed to the economy of the region, the safety of the inhabitants and the protection of natural values. What was less strongly developed, however, was a clear, concrete and widely accepted perspective (a policy or strategy) on the long-term future water resources development in the Mississippi Basin as a whole.

Studies carried out by the Federal Interagency Floodplain Management Review Committee (known as the Galloway Committee), further supported by US Army Corps of Engineers (USACE) studies, all prompted by the flood, have contributed substantially to debates over future policies for river management. Various alternatives for floodplain management were evaluated, but detailed concrete proposals were not asked for and were not included within the scope of this floodplain management assessment.

Identifying alternative future river management strategies that could be implemented raises two questions:

- Should the river management strategy be based on a continuation of the historic practice of river engineering works, using reservoirs and levees to provide a certain flood safety level, in combination with dredging works for navigation? Or alternatively, should a future strategy focus more on the use of the floodplain and the river in such a way that environmental values are enhanced, while at the same time the potential flood damage is reduced?
- How and to what extent can the river and floodplain resources be further used by society for the development of the regional and national economy without compromising the river ecosystem?

Associated with the need for a strategy for river basin management, additional questions include:

- How can federal and state agencies be organized in such a way that river basin management can take place effectively?
- How can management and development planning be structured so that stakeholders (industry, environmental groups, farmers, urban population and others) are actively involved in the process?

## 3.1. General History

A sketch of some aspects of the history of water resources development in the United States may provide some context for what has happened in the Mississippi River.

For more than a century, the Federal Government has been involved in developing water projects for a variety of purposes including flood control, navigation, power generation, and irrigation and settlement of the West. Most major water projects, such as large dams and diversions, were constructed by government agencies – the Bureau of Reclamation or the U.S. Army Corps of Engineers. Traditionally, the Corps has built and maintained projects designed primarily for flood control, navigation and power generation, whereas Bureau projects were designed primarily to enhance storage capacity and provide reliable supplies of water for irrigation and some municipal and industrial uses.

The Corps operates hundreds of flood control, navigation, and multipurpose works throughout the country. While its navigational activities date back to foundation of the Federal Government, the Corps did not become involved in flood control until the mid-to-late 1800s, and then was primarily concerned only with studies and investigations. Although Congress authorized the Corps to construct levees throughout the country in 1917, the modern era of federal involvement in flood control did not get fully underway until the enactment of the Flood Control Act of 1936, when Congress declared that it was in the national interest to assist states with flood control measures and that this was a 'proper' federal activity. Prior to this time, flood control had been largely viewed as a local responsibility. In addition to its flood control responsibilities, the Corps continued to construct navigational improvements, and later expanded its activities to include construction of multipurpose water projects.

For decades, the Federal Government took a largely structural approach to flood control, floodplain management and water supply development. While the Corps was active in building dams and levees throughout the country, channelling meandering rivers and streams to move floodwaters quickly and efficiently 'out of harms way', the Bureau was building some of the largest water supply projects in the world. Thus, the Corps and the Bureau are substantially involved in the management of some of the country's largest river systems, including the Colorado, the Columbia, the Missouri and the Mississippi.

Project construction for all types of water works expanded greatly during the 1930s and 1940s and continued rapidly until the late 1960s and early 1970s. By the late 1960s, a combination of changing national priorities and local needs, increasing construction costs and the decreasing number of prime locations for water works all contributed to a decline in new construction of major water works nationwide. Water supply for traditional off-stream uses – such as public supply, domestic, commercial, industrial and agricultural uses – was increasingly in direct competition with a growing interest in allocating water to maintain or enhance in-stream uses such as recreation, scenic enjoyment, and fisheries and wildlife habitat.

During the 1970s, construction of new projects slowed to a handful of major works, culminating in the completion of the Tellico Dam project in Tennessee and the Tennessee Tombigbee Waterway through Alabama and Mississippi. These projects pitted conservation and environmental groups, as well as some fiscal conservatives, against the traditional water resources development community. New on the scene was the National Environmental Policy Act of 1970 (NEPA), which for the first time required an assessment of the environmental effects of federal projects and provided for more public scrutiny of such projects.

In 1978, President Carter announced that future federal water policy would focus on improving water resources management, constructing only projects that were economically viable, cooperating with state and local entities, and sustaining environmental quality. The subsequent Reagan administration continued to oppose large projects, contending they were fiscally unsound. Federal water research and planning activities were also reduced during the early years of the Reagan Administration (early 1980s), which felt that states should have a greater role in carrying out such activities. Consistent with this outlook, President Reagan abolished the Water Resources Council, an umbrella agency established in 1968 to coordinate federal water policy and to assess the status of the nation's water resource and development needs.

Congress subsequently scaled back several remaining authorized projects, changed repayment and cost-share structures, and passed laws that altered project operations and water delivery programmes. For example, in 1982 Congress passed the Reclamation Reform Act, which altered the Bureau's water pricing policies for some users. The Act revised acreage limitation requirements and charges for water received to irrigate leased lands. Congress soon passed another landmark law, the Water Resources Development Act of 1986, which requires local entities to share the construction costs of water resources projects built by the Corps. This act did little to enhance comprehensive systems-wide planning but did allow projects to be built, especially if members of Congress wanted them.

Over the last decade, both the Corps and the Bureau have undertaken projects or programmes aimed at mitigating or preventing environmental degradation due in part to the construction and operation of large water projects. The agencies have pursued these actions through administrative efforts and congressional mandates, as well as in response to court actions. Currently, the Federal Government is involved in several restoration initiatives, including the Florida Everglades, the California Bay-Delta, the Mississippi Delta, Lake Ontario and St. Lawrence River, and the Columbia and Snake River Basins in the Pacific Northwest. Degradation processes in these river, lake and coastal systems have been occurring over many years.

These restoration initiatives are not without controversy. Each involves many stakeholders at the local and regional level (water users, landowners, farmers, commercial and sports fishermen, urban water suppliers and users, navigational interests, hydropower customers and providers, tourists and environmentalists), all with their own objectives and desires. At the same time, demand for traditional or new water resources projects continues – particularly for ways to augment local water supplies, maintain or improve navigation, and control or prevent floods and shoreline erosion.

## Now let us turn to the Mississippi

Nature's design of the upper Mississippi was imperfect for human use. So Congress, with the help of the US Army Corps of Engineers, remade it. For over a century, the Corps has responded to Congressional requests to provide for transportation and flood control – and most recently for ecology as well – in the upper Mississippi.

**Figure D.18. Barged cargo passing through a lock on the upper Mississippi River.**


**Figure D.19. The upper Mississippi River locks and navigation dams, known locally as pools.**

Nature's upper Mississippi was often less than four feet deep. Today's canalized version handles the standard upper-Mississippi barge: 2.75 m (9 ft) deep, 10.7 m (35 ft) wide and some 60 m (195 ft) long. They are tied three across (32 m or 105 ft) and five deep (297 m or 975 ft), and pushed (not towed) by a towboat, as shown in Figure D.18. Most of the locks are 33.5 m (110 ft) wide and 183 m (600 ft) long. So a standard set of barges is divided into two lengths, each taken separately through the lock, and then re-tied. A single lock-passage takes about twenty minutes. Waiting, because of traffic, can sometimes take twenty hours. This is one of the reasons why the Corps of Engineers has been asked to build more lock capacity.

Managing floods on the upper Mississippi, to the extent possible, is done by using reservoirs and levees. They will no doubt continue to be used.

The upper Mississippi has twenty flood-control reservoirs, controlled by three Corps of Engineer Districts: St Paul (16), Rock Island (3), and St Louis (1). There are also more than a hundred smaller dams and reservoirs on upper-Mississippi tributaries for local use and not for flood control on the main river. There are no flood control reservoirs on the river; the locks and dams on the river above St. Louis control water levels primarily for transportation (Figure D.19).

The Corps-controlled reservoirs serves several purposes. At each, a permanent minimum level is kept for recreation and wildlife. In dry times, water is released to maintain minimum levels in the tributary and the river. On a tributary that can flood any time, water is released after a flood, to be ready for the next one. On a tributary with well-defined flood seasons, floodwater is held for release as needed by the tributary and the river. Hydropower is produced at some reservoirs, and provides approximately 9% of the combined energy used in the Mid-continent Area Power Pool, which includes Iowa, Minnesota, Nebraska, North Dakota, South Dakota, and portions of Illinois, Montana and Wisconsin.

Levees are used for local flood control. They reduce flooding for one area but typically cause problems elsewhere. A levee on one side causes higher floods on the other. Levees on both sides narrow the channel, causing higher floods upstream. Levees protecting portions of the fertile floodplain keep floodwaters from occupying and soaking into that natural floodplain land. This increases the flood flow downstream.

The overall plan of reservoirs and levees includes the possibility that, in extreme situations, less valuable areas will be flooded to reduce the risk of losses in more valuable ones. According to the basic plan for the upper Mississippi, submitted to Congress by the Corps in 1940:

> The reservoir system proposed is not of sufficient capacity to completely control floods in the upper Mississippi River.... The best method of operation of reservoirs ... often requires that reservoirs be emptied as rapidly as practicable after a flood has passed in order to be in readiness for a possible second large flood in the same season.... Large discharges might be superimposed upon the crest of a major flood at some point downstream.

Hydrological regimes in watersheds have been changed significantly by the construction of dams, levees and channels. Drainage systems and other land use changes have influenced the runoff pattern. Watershed alterations promoted human welfare, and policy-makers saw opportunities to use water development as an engine for economic prosperity. The achievements were impressive: improved safety against floods, expanded navigation on the rivers, availability of floodplains for agricultural and commercial uses, production of hydroelectric power and other benefits.

However, as in many rivers that are highly controlled, the floodplains of the upper Mississippi, Illinois and Missouri Rivers have been reduced in size by the construction of levees. The remaining floodplains have slowly been rising due to silt deposition. River runoff has increased due to the loss of upland cover in the basin. The construction of wing dykes has made the low flow channel narrower and deeper. Plants and animal species have slowly disappeared. Wetland areas have decreased gradually as land was converted to agricultural use. However, even today the upper Mississippi, Illinois and Missouri Rivers still contain extensive and important ecological and landscape values. Figure D.20 illustrates land cover changes over a century at two sites along the upper Mississippi.

Comparison of the land-cover/land use data between 1891 and 1989 in the dammed portion of the upper Mississippi River showed that open water and marsh habitats generally increased, mostly at the expense of forest and agriculture (upper images in Figure D.20). Where river dykes have been built, wetland and woody areas tended to be converted to agricultural areas over the same period (lower images in Figure D.20).

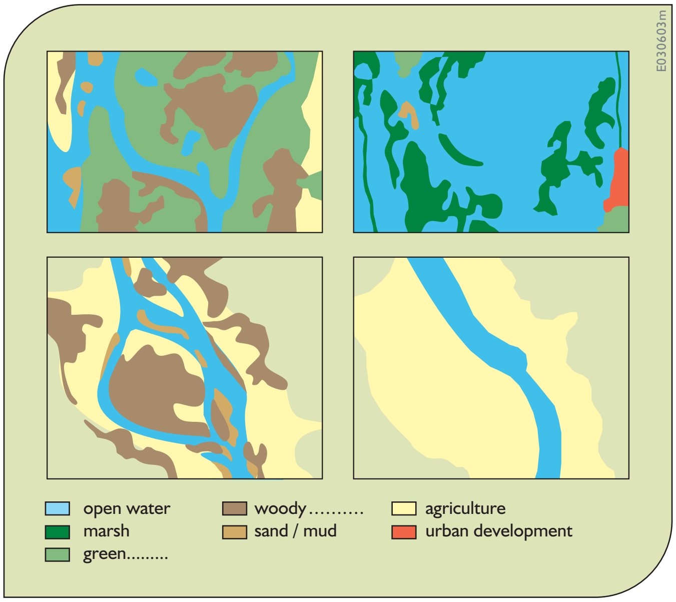



**Figure D.20.** Land-cover/land use comparisons at two sites on the upper Mississippi. The left-hand images depict land uses in 1891 and the right-hand images show the land uses at the same sites nearly 100 years later (Wlosinski et al., 2002.).

In all, nearly 2,000 km (1,240 miles) of levees now isolate more than 400,000 ha (988,000 acres) from the river during all but the highest discharge rates. Wing dams and levees, along with other changes to the watershed, have also had a major effect on habitats by changing the relationship between discharge and water-surface elevations. Wing dams have narrowed and deepened the main channel so that water elevations at low discharges are now lower than they were historically. Levees restrict flows and result in higher water elevations during high discharges. Water surface elevations at relatively low discharges (1,700 $m^3$/s or 60,000 $ft^3$/s) have dropped about 2.4 m (8.0 ft) over the recorded 133-year period at St. Louis, Missouri, 0.5 m (1.5 ft) over the 52-year record at Chester, Illinois, and 1.5 m (5.0 ft) over the 60-year record at Thebes, Illinois. Water-surface elevations at relatively high discharges (22,000 $m^3$/s or 780,000 $ft^3$/s), however, have risen about 2.7 m (9 ft) over the record period at St. Louis, 1.5 m (5.0 ft) at Chester, and 1.1 m (3.6 ft) at Thebes.

In the upper Mississippi Basin, protection levels against flooding vary, largely depending on the use of the floodplain. Major urban areas are generally protected by levees that are strong and high enough to withhold a flood like the 1993 flood. Many small towns and agricultural areas, however, are protected by 50-year levees, and many of these were overtopped or failed in 1993.

The 1993 flood reminded everyone of how damaging a river can be. The damage caused by this 1993 flood was substantial but was less than it might have been without the upstream flood control works (reservoirs and levees), upland soil conservation measures, terraces and ponds. Yet those flood damage reduction measures were not sufficient. The issue addressed by Gilbert White (1945) over a half century ago, the issue today, and no doubt the issue on into the future is how to effectively reduce the threat of future flood damage while supporting continued economic development in the affected areas.

A variety of options are available to reduce flood risks. Many are similar to those being considered in the Netherlands. They include raising or strengthening levees, relocating or setting back levees, dredging the main and/or side channels, and lowering floodplains. These may have major or minor effects, be expensive or relatively inexpensive, be in line or out of line with the current regulations or policies, and be more or less socially acceptable. For sustainable protection against flooding, measures in the upstream and downstream part of the river need to be considered together as a single system.

### 3.2. Other Considerations

The Mississippi is an important waterway for inland navigation, and the region's economy depends on the inexpensive transport of its bulk commodities. Much of the transport of goods to and from the Midwest states is by barge through an extensive lock and dam system. Periodic dredging is necessary at several locations on the river to maintain the navigation system.

Safe, speedy and cheap navigation demand a channel of adequate depth and width, and locks large enough to accommodate the fifteen-barge tows (Figure D.18). For such tows, the length of the present locks on the upper Mississippi River and Illinois is limiting, and often causes considerable delays, as mentioned above. Furthermore, the continued growth of commercial and recreational traffic is approaching maximum lock capacities. The width of the navigation channel is insufficient for two-way navigation at many locations. This further delays navigation and adds to transportation costs. However, reliability and low cost still make inland navigation a competitive means of transport. Other positive attributes of inland navigation are the relatively low energy consumption and the large transport capacity with few nuisance and safety problems. Air pollution by navigation per ton of cargo is far less than that produced per ton by trucks, or even by rail transport.

Agriculture and industry compete for space on the floodplain. Agriculture is the leading commercial user of the floodplains along the rivers in the upper Mississippi Basin. About 70% of the total 2.4 million ha (6 million acres) of floodplain is used for agricultural production. Due to the fertile soils, the yield in well-drained floodplains is usually substantially higher than in the upland areas.

A small but highly valuable portion of the floodplain, mainly in the direct vicinity of towns, is occupied by industries. Floodplains well protected by levees offer industry the advantage of immediate access to inland water transport to bring in raw materials and transport finished products, and immediate access to cooling and process water.

Nature also competes for space on the floodplain. Large wetland areas used to exist in the basin. These are unique links between land and water. Some are almost continuously under water, whereas others may be flooded for only a short period. This results in a variety of wetlands and their specific habitat types and functions. Through these special conditions, wetlands are among the most biologically productive natural ecosystems in the world.

In addition to being vital to the survival of various animals and plants, including threatened and endangered species, wetlands play a role in the reduction of peak water levels during smaller flood events (occuring every few years) because of their capability to store floodwater and release it slowly. For large events like the 1993 flood, the effect of wetlands on high water levels is much less because wetlands are usually saturated by rainfall or already flooded in an earlier stage.

From the mid-1800s the wetland areas were gradually decreased mainly to reduce the incidence of malaria and other waterborne diseases. Wetlands were converted to agricultural land and to a lesser extent to residential and industrial uses. Currently, wetlands account for about 10% of the floodplain in the upper Mississippi Basin.

Wetlands, storage lakes and the river itself provide considerable opportunities for recreation. Popular activities are hunting, fishing, camping, boating, sightseeing and bird watching. River-related recreation is important for the economy of local communities. The locks and dams provide many recreation lakes and the opportunity for boating, skiing, fishing and hunting.

Floodplains along the rivers also contain numerous archaeological and historic sites. These sites include historic architectural and engineering features and structures, and resources of traditional cultural or heritage significance to Native Americans and other social or cultural groups. Examples are forts, quarries, potteries and burial sites. Construction activities in and along the river, streambank erosion and extreme floods all can adversely affect these values.

Some of these multiple uses of the floodplain are compatible with flood management and obviously others are not. Yet all must be considered when developing a flood management plan.

The river itself is a source of water for municipal and industrial use: drinking water, processing and cooling water, and irrigation. The number of people who obtain drinking water from the river and its tributaries is quite large, yet the amount of water extracted for these purposes is relatively small when compared with the river discharge. As long as the flow withdrawn for water uses (after proper treatment) is returned to the rivers, this water use can be considered of minor importance in the overall water balance of the rivers. The export of water to other basins (diverting Missouri water all the way to Denver, for example) is another matter.

## 3.3. Interactions Among User Groups

Just as there are multiple purposes for the upper Mississippi River water and floodplain, some of which are complementary and others of which conflict, the same is true of the different user groups or stakeholders.

Table D.1 provides a rough indication of the ranges of some of the potentially more conflict-causing interactions. The table's columns are the activities that affect the activities shown in the rows. Scores of +1 and +2 indicate complementarities and positive interactions, 0 is neutral and –1 and –2 are competing or negative interactions. Participatory planning can help to push these relative numbers towards the positive side of the ranges, representing fewer conflicts and more positive interactions.

There are other important uses of water that are not mentioned in Table D.1 because the interactions are obvious and straightforward. Supplying municipal and domestic public water systems is one. Hydropower is another.

Some of the categories of conflict indicated in Table D.1 create serious problems in the upper Mississippi Basin. Some of the potential positive interactions are only weakly developed. The sharpest conflicts appear to be those between environmentalists and agriculturalists, and with those who would develop industrial sites through flood control and increased river shipping, but are constrained by legislation limiting land development in the floodplain.

A significant number of farmers consider environmental concerns by-and-large as expressions of urban outsiders, who have somehow gained the right to criticize the farmers' activities on their own land. In some cases, environmental regulations are viewed as threats to farm livelihood. Furthermore, when the reasons for regulation involve the preservation of animal or plant species that seem either insignificant or even nuisances to the agricultural community, the regulations are perceived to be both insulting and injurious. Whatever the pros and cons of the regulations, when the conflict of interests has reached the stage that some of the parties feel threatened, willing compliance with regulations drops and conflict resolution becomes far more difficult.

Table D.1 indicates only direct interactions. Indirect interactions are not listed, although they may also be very important. For instance, the effect of industrial development in or near the floodplain on river ecology is scored at 0 to –2 because industrial effects may vary between environmentally neutral, moderately polluting to so badly polluting that they destroy habitats. Industry is rarely directly complementary or positive to environmental concerns in floodplains. However, if the commitment is made to use some of the revenues generated by industrial development, through taxes or other means, to enhance environmental conditions (e.g. using some funds to establish conservation areas) then the indirect effects of the industrial development on the environment may be positive (score +1). Such indirect effects can be important in negotiations among user groups.

E030603n

| | scoring the effects of | | | | | | |
|---|---|---|---|---|---|---|---|
| | flood control structures | agriculture in the floodplain | industry in or near floodplain | shipping channel dredging | river based recreation | historical and cultural preservation | environmental protection |
| **on flood control** | | + 1 to − 1 | 0 | + 1 to − 1 | 0 to − 1 | + 1 to − 1 | + 1 to − 2 |
| **on floodplain agriculture** | + 2 to + 1 | | 0 to − 1 | + 2 to + 1 | 0 to − 1 | + 1 to − 1 | + 1 to − 2 |
| **on industry in or near floodplain** | + 2 | + 1 to 0 | | + 2 to + 1 | 0 | 0 to − 1 | 0 to − 2 |
| **on river navigation** | − 1 to + 1 | 0 | + 1 to 0 | | 0 to − 2 | 0 | 0 to − 1 |
| **on river-based recreation** | + 1 to − 1 | 0 to − 1 | 0 to − 2 | 0 to − 1 | | + 1 to − 1 | + 2 to + 1 |
| **on historical & cultural property** | + 2 to − 2 | 0 to − 1 | 0 to − 2 | 0 to − 1 | + 1 to 0 | | + 2 to 0 |
| **on condition of river ecology** | 0 to − 2 | + 1 to − 1 | 0 to − 2 | 0 to − 1 | + 1 to − 1 | + 2 to 0 | |

+ 2 = necessary, or highly complementary and positive effects
+ 1 = generally positive
0 = neutral: no particular advantage nor conflict (small interaction)
− 1 = implementation of one means a restriction or damage to the other
− 2 = mutually exclusive thus completely in competition, or highly damaging

**Table D.1.** Interactions among potentially conflicting river activities.

There are serious conflicts over the degree and extent of flood control that should be provided and the ways this should be done. There are various possibilities between two extremes: those who feel that flood protection levels should be raised significantly, primarily by increasing the heights of levees all along the river; and those who believe that development in the floodplain should be actively discouraged, and existing uses that restrict flood flows should be reduced. Arguments for the first option are largely that economic development in the region, particularly farming and industry, will be greatly served by protection from floods. The arguments for the second approach are primarily that floodplain protection and compensation for flood damage are very expensive, and that there is no compelling reason for the nation to subsidize floodplain development when alternative sites for agriculture and industry exist outside the floodplain. Floodplain protection measures are also frequently not environmentally friendly. Decreased floodplain development allows a return to more natural conditions.

Both arguments are valid. In some cases, flood control measures are complementary to environmental interests, particularly when the measures involve setting back levees, lowering floodplains or making parallel, unregulated channels or levees that protect or encourage the re-establishment of certain kinds of habitat. In general, however, environmental interests favour less engineering and lower rather than higher levees. River dredging is highly controversial. This puts environmental groups in strong conflict with those advocating floodplain development.

Negotiating conflicts between different interest groups requires extensive data collection and expert analyses of the impacts of different measures to control flooding. The results of the analyses need to be presented in ways that are accessible and understandable to stakeholders and decision-makers, and there should be a process that encourages debate, negotiation and compromises among stakeholders.

### 3.4. Creating a Flood Management Strategy

The current (2004) approach to river management on the upper and middle Mississippi River is not as integrated or as well balanced as it could be. This situation is not unique to this river and some of the reasons for it could be eliminated if policies that restrict comprehensive integrated approaches were changed. The 1993 flood in the upper Mississippi River Basin focused national attention on this problem, at least for a while. The concern it aroused presented an opportunity for reassessment and the creation of an integrated water resources management strategy. One of the conclusions of the Galloway Committee (1994) was: '*The United States has a rare opportunity to make a change in floodplain management. It should not be missed*'. Many of the recommendations made by this Committee have yet to be seriously considered. It seems, once again, that the concerns generated by the flood and identified by the committee are fading from national attention. What began as a useful debate following the 1993 flood seems to be dissipating without many substantial decisions that will make it easier to manage or mitigate the damage when the next flood occurs.

An argument could be made that the resources offered by the river could be used considerably more intensively than they are now without compromising the river's ecological integrity. The basin's potential for increased economic and ecological benefits is not fully developed. A more active management of ecological values, including using options for ecological restoration, can serve environmental protection and at the same time allow sustainable economic development. Active management of both economic and ecological values requires some compromises, but can also provide some important gains.

The economic advantage the river offers to the communities along it and to the region could and should be increased along with its ecological benefits. Navigation could be increased, even from an environmental point of view, given the relatively low air pollution and high safety levels of river freight compared with trucking and even rail. River transport of containerized cargo in addition to bulk agricultural products, agro-chemicals, coal and steel could be stimulated. There is potential to develop additional river ports and inter-modal transport facilities. Such development would enhance the economies of the communities along the river.

Similarly, the use of the river and the riverbanks for recreation (boating, fishing, hunting) could be increased. This would benefit local economies and could lead to the provision of guarantees for environmental conservation. Increased river-related recreation helps support environmental integrity since a high-quality environment is a necessary condition for this activity.

An integrated water management strategy would address flood management issues and uncertainties such as how levees influence upstream flood stages, how raising or setting back levees affect water levels elsewhere, how floodplain vegetation (particularly trees) influences flood stages, and so on. In addition, the flood safety levels for different land uses on the floodplain would be selected. The Galloway Report (1994) emphasized flexibility and variations in flood protection levels. Questions remain as to what levels of flood protection are appropriate for urban areas/industrial sites and agricultural lands, and how agreements on such levels are to be reached.

### 3.5. The Role of the Government and NGOs

State and federal agencies are major players in water resources management in this as in most major river basins in the United States. Without their leadership, not much planning or management takes place or is implemented. Yet there are limits as to what any particular agency can do. Thus, it is absolutely necessary for them to work together with each other and with private non-governmental organizations (NGOs and other stakeholder groups) if any approximation of an integrated water management plan or strategy is to be developed. If institutional change can facilitate that process, such change should be considered. Independent river basin commissions representing all stakeholders and taking responsibility for integrated river management can be effective, assuming their recommendations are sought and would be seriously considered in the higher-level political decision-making process.

Present regulations regarding benefit–cost analysis need to be revised to allow more types of values to be considered alongside monetary values. The nation's interest in the economic development of the upper Mississippi, Illinois and Missouri River Basins is not reflected in the restricted way future benefits are calculated in the benefit–cost considerations required for federal investment plans in the river and floodplains. The omission of any estimates of the growth of future economic benefits can undervalue economic development. Since environmental benefits are not easily translated into monetary terms, they tend not to be fully considered. The economic benefits derived from natural well-functioning ecosystems that help to justify environmental protection are not made explicit either. The manner in which future benefits are included in the benefit–cost calculation should be consistent for different types of projects.

A balanced river-basin development plan for the upper Mississippi, Illinois and Missouri River Basins, to be drafted interactively with wide participation by all parties concerned, would improve understanding of the complex interrelationships between environmental protection, resource use, and river and floodplain development. Such a development plan would not be static, but would serve as a foundation to be revised on a regular basis in line with new developments and insights. In such an adaptive way, the plan could enhance the effective as well as efficient use of the river resources.

# 4. Flood Risk Reduction

There are two types of structural alternatives for flood risk reduction. One is flood storage capacity in reservoirs designed to reduce downstream peak flood flows. The other is channel enhancement and/or flood-proofing works that will contain peak flood flows and reduce damage. This section introduces methods of modelling both these alternatives for inclusion in either benefit–cost or cost-effectiveness analyses. The latter analyses apply to situations in which a significant portion of the flood control benefits cannot be expressed in monetary terms and the aim is to provide a specified level of flood protection at minimum cost.

The discussion will first focus on the estimation of flood storage capacity in a single reservoir upstream of a potential flood damage site. This analysis will then be expanded to include downstream channel capacity improvements. Each of the modelling methods discussed will be appropriate for inclusion in multipurpose river basin planning (optimization) models having longer time-step durations than those required to predict flood peak flows.

## 4.1. Reservoir Flood Storage Capacity

In addition to the active storage capacities in a reservoir, some capacity may be allocated for the temporary storage of flood flows during certain periods in the year. Flood flows usually occur over time intervals lasting from a few hours up to a few days or weeks. Computational limitations make it impractical to include such short time durations in many of our multipurpose planning models. If we modelled these short durations, flood routing equations would have to be included in the model; a simple mass balance would not be sufficient. Nevertheless, there are ways of including unknown flood storage capacity variables within longer-period optimization models. These capacities can be determined on the basis of economic efficiency within benefit–cost analyses, or they can be based on constraints specifying maximum flood risk protection levels.

Consider a potential flood damage site along a river. A flood control reservoir can be built upstream of that site. The question is how much flood storage capacity, if any, should it contain. For various assumed capacities, simulation models can be used to predict the impact on the downstream flood peaks. These hydraulic simulation models must include flood routing procedures from the reservoir to the downstream potential damage site and the flood control operating policy at the reservoir. For various downstream flood peaks, economic flood damages can be estimated. To calculate the expected annual damage associated with any upstream reservoir capacity, the probability of various damage levels being exceeded in any year needs to be calculated.

The likelihood of a flood peak of a given magnitude or greater is often defined by its expected return period. How many years would one expect to wait, on average, to observe another flood equal to or greater than some specified magnitude? This is the reciprocal of the probability of observing such a flood in any given year. A $T$-year flood has a probability of being equalled or exceeded in any year of $1/T$. This is the probability that could be calculated by adding up the number of years an annual flood of a given or greater magnitude is observed, say in 1,000 or 10,000 years, divided by 1,000 or 10,000 respectively. A one-hundred year flood or greater has a probability of 1/100 or 0.01 of occurring in any given year. Assuming annual floods are independent, if a 100-year flood occurs this year, the probability that a flood of that magnitude or greater occurring next year remains 1/100 or 0.01.

If $PQ$ is the random annual peak flood flow and $PQ_T$ is a particular peak flood flow having a return period of $T$ years, then by definition the probability of an actual flood of $PQ$ equalling or exceeding $PQ_T$ is $1/T$;

$$\Pr[PQ \geq PQ_T] = 1/T \tag{D.1}$$

The higher the return period (i.e. the more severe the flood), the lower the probability that a flood of that magnitude or greater will occur. Equation D.1 is plotted in Figure D.21.

The exceedance probability distribution shown in Figure D.21 is simply 1 minus the cumulative distribution function $F_{PQ}(\cdot)$ of annual peak flood flows. The area under the function is the mean annual peak flood flow, E[$PQ$].

The expected annual flood damage at a potential flood damage site can be estimated from an exceedance

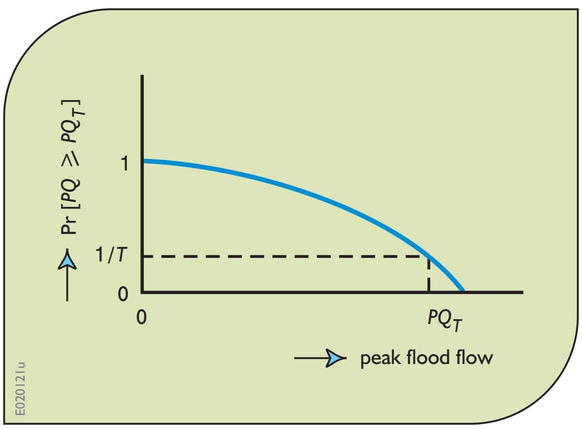



**Figure D.21. Probability of annual peak flood flows being exceeded.**

probability distribution of peak flood flows at that site together with a flow or stage damage function. The peak flow exceedance distribution at any potential damage site will be a function of the upstream reservoir flood storage capacity $K_f$ and the reservoir operating policy.

The probability that flood damage of $FD_T$ associated with a flood of return period $T$ will be exceeded is precisely the same as the probability that the peak flow $PQ_T$ that causes the damage will be exceeded. Letting $FD$ be a random flood damage variable, its probability of exceedance is

$$\Pr[FD \geq FD_T] = 1/T \tag{D.2}$$

The area under this exceedance probability distribution is the expected annual flood damage, $E[FD]$;

$$E[FD] = \int_0^\infty \Pr[FD \geq FD_T]\mathrm{d}FD_T \tag{D.3}$$

This computational process is illustrated graphically in Figure D.22. The analysis requires three input functions that are shown in quadrants (a), (b), and (c). The dashed-line rectangles define point values on the three input functions in quadrants (a), (b), and (c) and the corresponding probabilities of exceeding a given level of damages in the lower right quadrant (d). The distribution in quadrant (d) is defined by the intersections of these dashed-line rectangles. This distribution defines the probability of equalling or exceeding a specified level of damage in any given year. The (shaded) area under the derived function is the annual expected damage, $E[FD]$.

The relationships between flood stage and damage, and flood stage and peak flow, defined in quadrants (a) and (b) of Figure D.22, must be known. These do not depend on the flood storage capacity in an upstream reservoir. The information in quadrant (c) is similar to that shown in Figure D.21 defining the exceedance probabilities of each peak flow. Unlike the other three functions, this distribution depends on the upstream flood storage capacity and flood flow release policy. This peak flow probability of exceedance distribution is determined by simulating the annual floods entering the upstream reservoir in the years of record.

The difference between the expected annual flood damage without any upstream flood storage capacity and the expected annual flood damage associated with a flood storage capacity of $K_f$ is the expected annual flood damage reduction. This is illustrated in Figure D.23.

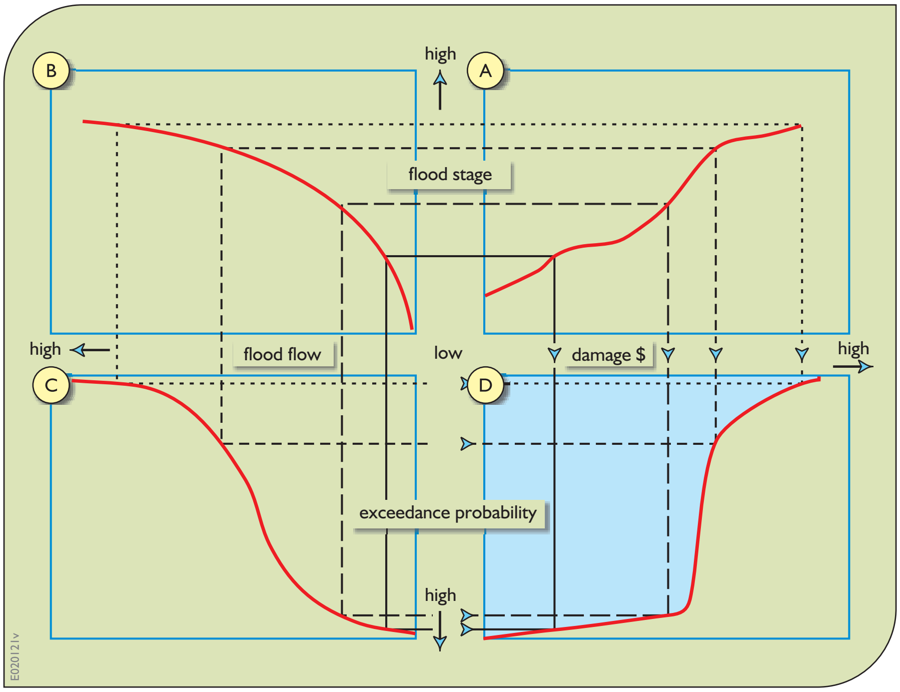



**Figure D.22.** Calculation of the expected annual flood damage shown as the shaded area in quadrant (d) derived from the expected stage-damage function (a), the expected stage-flow relation (b), and the expected probability of exceeding an annual peak flow (c).

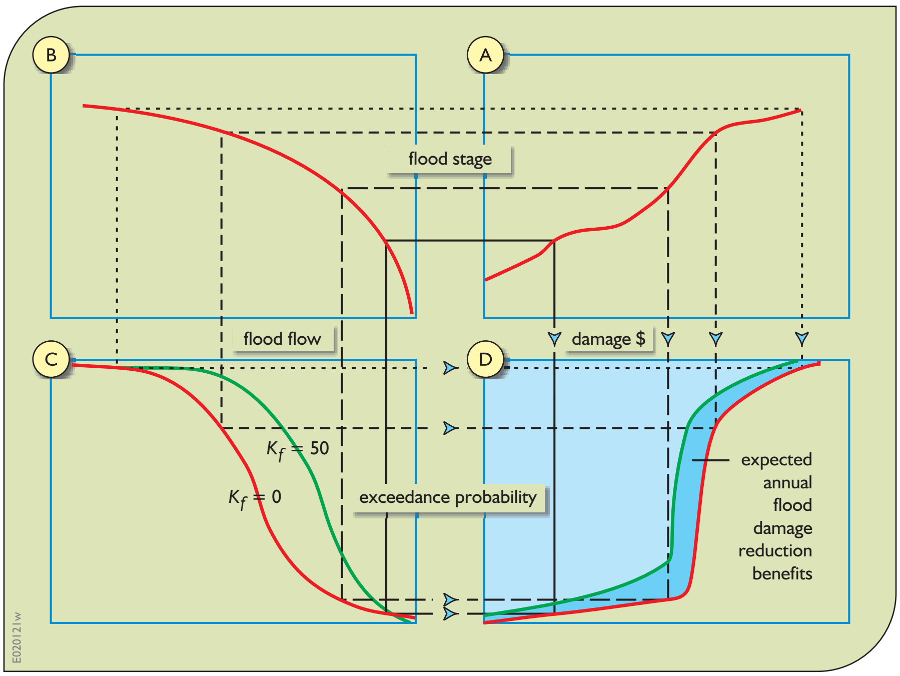



**Figure D.23.** Calculation of expected annual flood damage reduction benefits, shown as the darkened portion of quadrant (d), associated with a specified reservoir flood storage capacity.

Knowing the expected annual damage reduction associated with various flood storage capacities, $K_f$, permits the definition of a flood damage reduction function, $B_f(K_f)$.

If the reservoir is a single-purpose flood control reservoir, the eventual tradeoff is between the expected flood reduction benefits, $B_f(K_f)$, and the annual costs, $C(K_f)$, of that upstream reservoir capacity. The particular reservoir flood storage capacity that maximizes the net benefits, $B_f(K_f) - C(K_f)$, may be appropriate from a national economic efficiency perspective but may not be best from a local perspective. Those occupying the potential damage site may prefer a specified level of protection from the reservoir storage capacity, rather than the protection that maximizes expected annual net benefits, $B_f(K_f) - C(K_f)$.

If the upstream reservoir is to serve multiple purposes, say for water supply, hydropower and recreation as well as for flood control, the expected flood reduction benefit function just derived could be a component in the overall objective function for that reservoir.

Total reservoir capacity $K$ will equal the sum of dead storage capacity $K_d$, active storage capacity $K_a$ and flood storage capacity $K_f$, assuming they are the same in each period $t$. In some cases they may vary over the year. If the required active storage capacity can occupy the flood storage zone when flood protection is not needed, the total reservoir capacity $K$ will be the dead storage, $K_d$, plus the maximum of either 1) the actual storage volume and flood storage capacity in the flood season or 2) the actual storage volume in non-flood season.

$K \geq K_d + S_t + K_f$ for all periods $t$ in flood season plus the following period (that represents the end of the flood season) (D.4)

$K \geq K_d + S_t$ for all remaining periods $t$ (D.5)

In the above equations, the dead storage capacity, $K_d$, is a known variable. It is included in the capacity Equations D.4 and D.5, assuming that the active storage capacity is greater than zero. Clearly, if the active storage capacity were zero, there would be no need for dead storage.

## 4.2. Channel Capacity

The unregulated natural peak flow of a particular design flood at a potential flood damage site can be reduced by upstream reservoir flood storage capacity, or it can be contained within the channel at the potential damage site by levees and other channel-capacity improvements. In this section, the possibility of levees or dykes and other channel capacity or flood-proofing improvements at a downstream potential damage site will be considered. The approach used will provide a means of estimating combinations of flood control storage capacity in upstream reservoirs and downstream channel capacity improvements that together will provide a pre-specified level of flood protection at the downstream potential damage site.

Let $QN_T$ be the unregulated natural peak flow in the flood season having a return period of $T$ years. Assume that this peak flood flow is the design flood for which protection is desired. To protect from this design peak flow, a portion $QS$ of the peak flow may be reduced by upstream flood storage capacity. The remainder of the peak flow $QR$ must be contained within the channel. Hence, if the potential damage site $s$ is to be protected from a peak flow of $QN_T$, the peak flow reductions due to upstream storage, $QS$, and channel improvements, $QR$, must at least equal to that peak flow;

$$QN_T \leq QS + QR \tag{D.6}$$

The extent to which a specified upstream reservoir flood storage capacity reduces the design peak flow at the downstream potential damage site can be obtained by routing the design flood through the reservoir and the channel between the reservoir and the downstream site. Doing this for a number of reservoir flood storage capacities permits the definition of a peak flow reduction function, $f_T(K_f)$:

$$QS = f_T(K_f) \tag{D.7}$$

This function is dependent on the relative locations of the reservoir and the downstream potential damage site, on the characteristics and length of the channel between the reservoir and downstream site, on the reservoir flood control operating policy, and on the magnitude of the peak flood flow.

An objective function for evaluating these two structural flood control measures should include the cost of reservoir flood storage capacity, $\text{Cost}_K(K_f)$, and the cost of channel capacity improvements, $\text{Cost}_R(QR)$, required to contain a flood flow of $QR$. For a single-purpose, single-damage site, single-reservoir flood control problem, the minimum total cost required to protect the potential damage site from a design flood peak of $QN_T$, may be obtained by solving the model:

$$\text{minimize } \text{Cost}_K(K_f) + \text{Cost}_R(QR) \tag{D.8}$$

subject to

$$QN_T \leq f_T(K_f) + QR \tag{D.9}$$

Equations D.8 and D.9 assume that a decision will be made to provide protection from a design flood $QN_T$ of return period $T$; it is only a question of how to provide the required protection.

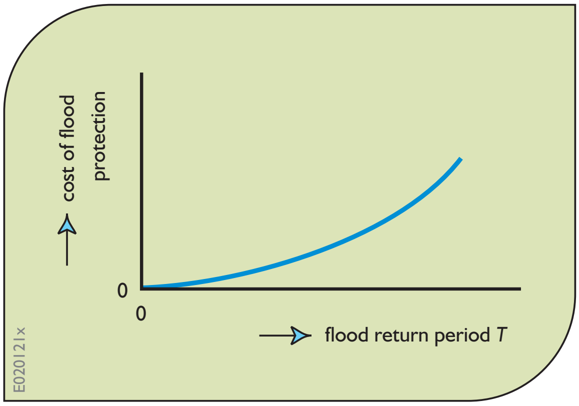



**Figure D.24. Tradeoff between minimum cost of flood protection and flood risk, as expressed by the expected return period.**

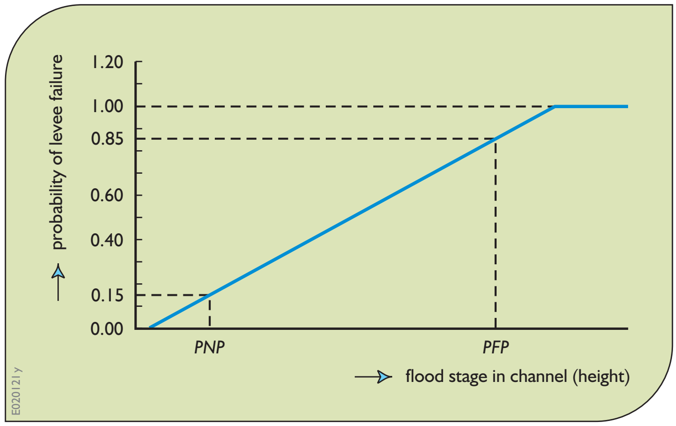



**Figure D.25. Assumed cumulative probability distribution of levee failure expressed as function of flood stage in river channel.**

Solving Equations D.8 and D.9 for peak flows $QN_T$ of various return periods $T$ will identify the risk–cost tradeoff. This tradeoff function might look like what is shown in Figure D.24.

## 4.3. Estimating Risk of Levee Failures

Levees are built to reduce the likelihood of flooding on the floodplain. Flood flows prevented from flowing over a floodplain due to a levee will have relatively little effect on users of the floodplain, unless of course the levee fails to contain the flow. If any of the flow in the stream or river channel passes over, through or under the levee and onto the floodplain, the levee is said to fail. This can result from flood events that exceed (overtop) the design capacity of the levee, or from various types of geo-structural weaknesses. The probability of levee failure along a river reach is in part a function of the levee height, the probability distribution of flood flows in the stream or river channel and the probability of geo-structural failure. The latter depends in part on how well the levee and its foundation were constructed. Some levees are purposely designed to 'fail' at certain sites at certain flood stages to reduce the likelihood of more substantial failures and flood damage further downstream.

The probability of levee failure given the flood stage (height) in the stream or river channel is often modelled using two flood stages. The US Army Corps of Engineers calls the lower stage the probable non-failure point, *PNP*, and the higher stage the probable failure point, *PFP* (USACE, 1991). At the *PNP*, the probability of failure is assumed to be 15%. Similarly, the probability of failure at the *PFP* is assumed to be 85%. A straight-line distribution between these two points is also assumed, as shown in Figure D.25. Of course, these points and distributions are at best only guesses, as not many, if any, data will exist to base them on at any given site.

To estimate the risk of a flood in the floodplain protected by a levee due to overtopping or geo-structural levee failure, the relationships between flood flows and flood stages in the channel and on the floodplain must be defined.

Assuming no geo-structural levee failure, the flood stage in the floodplain protected by a levee is a function of the flow in the stream or river channel, the cross-sectional area of the channel between the levees on either side, the channel slope and roughness, and the levee height. If floodwaters enter the floodplain, the resulting water level or stage in the floodplain will depend on the topological characteristics of the floodplain. Figure D.26 illustrates the relationship between the flood stage in the channel and the flood stage in the floodplain, assuming no geo-structural failure of the levee. Obviously once the flood begins overtopping the levee, the flood stage in the floodplain begins to increase. Once the flood flow is of sufficient magnitude that its stage without the levee is the same as that with the levee, the existence of a levee has only a negligible impact on the flood stage.

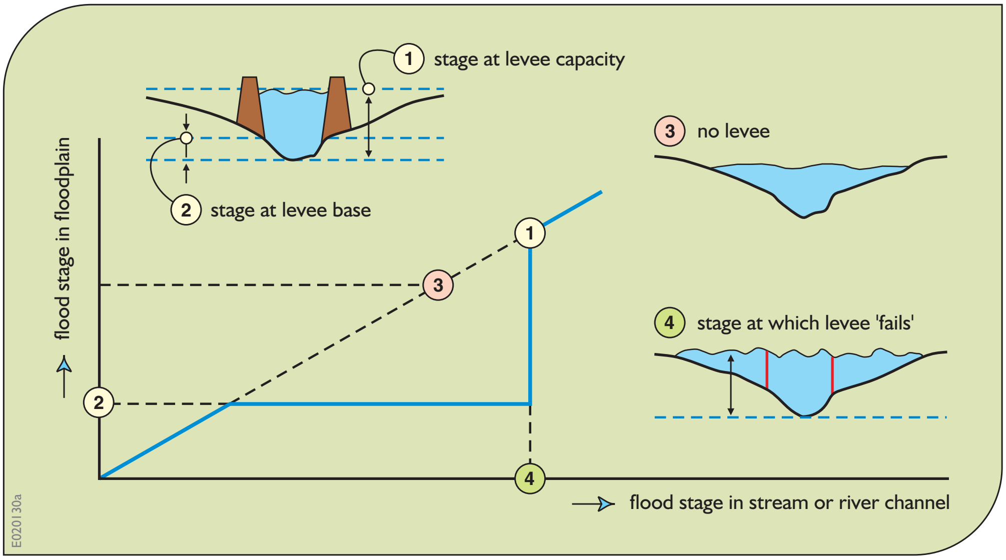



**Figure D.26.** Influence of a levee on the flood stage in floodplain compared to stream or river flood stage. The channel flood stage where the curve is vertical is the stage at which the levee fails due to overtopping or from geo-structural causes.

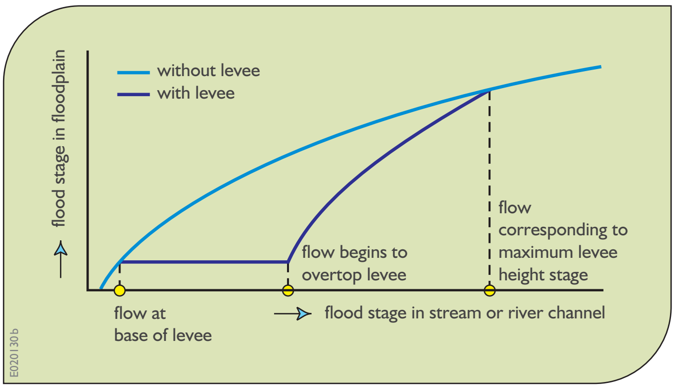



**Figure D.27.** Relationship between flood flow and flood stage in a floodplain with and without flood levees, again assuming no geo-structural levee failure.

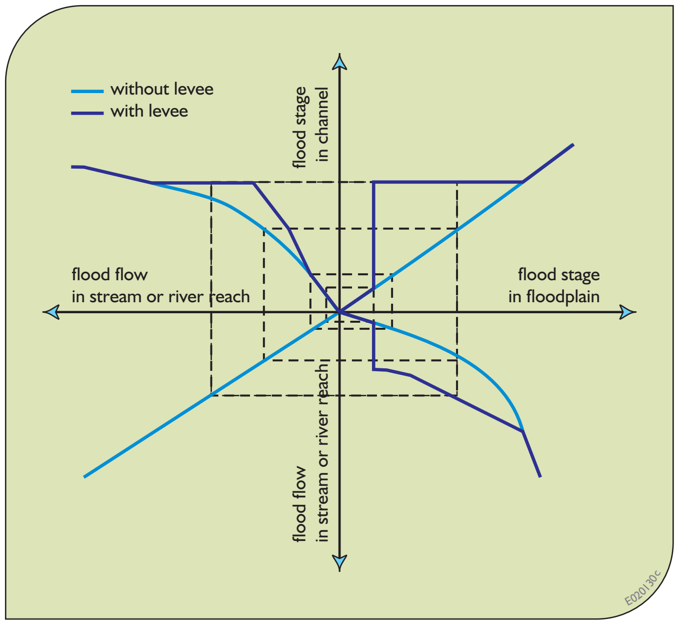



**Figure D.28.** Deriving the relation (shown in the upper left quadrant) between flood flow and flood stage in the channel.

Figure D.27 illustrates the relationship between flood flow and flood stage in a floodplain with and without flood levees, again assuming no geo-structural levee failure.

Combining Figures D.26 and D.27 defines the relationship between reach flow and channel stage. This is illustrated in the upper-left quadrant of Figure D.28.

Combining the relationship between flood flow and flood stage in the channel (upper-left quadrant of Figure D.28) with the probability distribution of levee failure (Figure D.25) and the probability distribution of annual peak flows being equalled or exceeded (Figure D.21), provides an estimate of the expected probability of levee failure. Figure D.29 illustrates this process of calculating the expected probability of levee failure in any year from overtopping and/or geo-structural failure. This expected probability is shown as the blue shaded area in the lower-right quadrant of the figure.

The channel flood stage function, $S(q)$, of peak flow $q$ shown in the upper-left quadrant of Figure D.29 is obtained from the upper-left quadrant of Figure D.28. The probability of levee failure, $PLF(S)$, a function of

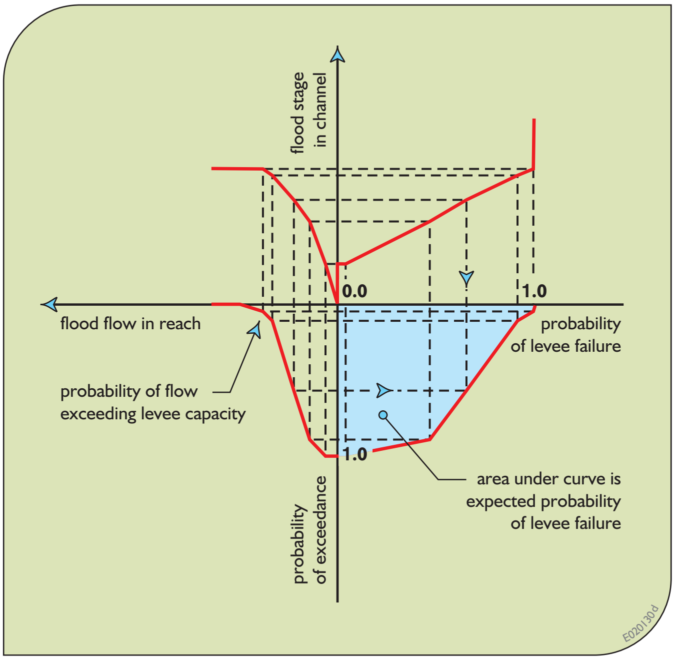



Figure D.29. Derivation of the probability of exceeding a given probability of levee failure, shown in lower right quadrant. The shaded area enclosed by this probability distribution is the annual expected probability of levee failure.

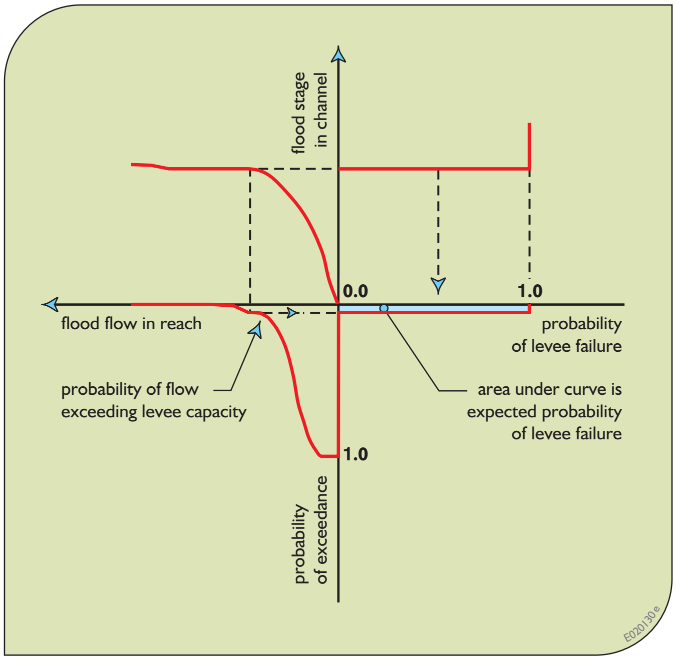



Figure D.30. Calculation of annual probability of equalling or exceeding any specified probability of levee overtopping, shown in the lower right quadrant. The shaded area in this quadrant is the expected probability of levee overtopping, assuming there is no geo-structural failure.

flood stage, $S(q)$, shown in the upper-right quadrant is the same as in Figure D.25. The annual peak flow exceedance probability distribution, $F_Q(q)$, (or its inverse $Q(p)$) in the lower-left quadrant is the same as Figure D.30 or that in the lower-left quadrant (c) of Figure D.22. The exceedance probability function in the lower-right quadrant of Figure D.29 is derived from each of the other three functions, as indicated by the arrows, in the same manner as described in Figure D.22.

In mathematical terms, the annual expected probability of levee failure, E[*PLF*], found in the lower-right quadrant of Figure D.29, equals

$$\begin{aligned} \mathrm{E}[PLF] &= \int_0^\infty PLF[S(q)]f(q)\mathrm{d}q = \int_0^1 PLF[S(Q(p))]\mathrm{d}p \\ &= \int_0^1 PLF'(p)\mathrm{d}p \qquad \text{(D.10)} \end{aligned}$$

where $PLF'(p)$ is the probability of levee failure associated with a flood stage of $S(q)$ having an exceedance probability of $p$.

Note that if the failure of the levee was only due to channel flood stages exceeding the levee height (i.e. if the probability of geo-structural failure were zero), the expected probability of levee failure would be simply the probability of exceeding a channel flow whose stage equals the levee height, as defined in the lower-left quadrant of Figure D.29. This is shown in Figure D.30.

Referring to Figure D.30, if the levee height is increased, the horizontal part of the curve in the upper-right quadrant would rise, as would the horizontal part of the curve in the upper-left quadrant as it shifts to the left. Hence, given the same probability distribution as defined in the lower-left quadrant, the expected probability of exceeding an increased levee capacity would decrease, as it should.

## 4.4. Annual Expected Damage From Levee Failure

A similar analysis can provide an estimate of the expected annual floodplain damage for a stream or river reach. Consider, for example, a parcel of land on a floodplain at some location $i$. If an economic efficiency objective were to guide the development and use of this parcel, then the owner would want to maximize the net annual economic benefits derived from its use, $B_i$, less the annual (non-flood damage) costs, $C_i$, and the expected annual

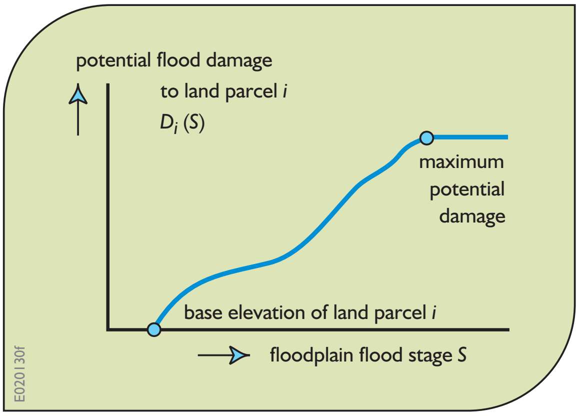



**Figure D.31. An example function defining the damages that will occur given any flood stage *S* to a parcel of land *l*.**

flood damage, $EAD_i$. The issue of concern here is the estimation of the expected annual flood damage.

The degree of damage at location $i$ resulting from a flood will depend in part on the depth of flooding at that location and a host of other factors (flood duration, velocity of and debris in flood flow, time of year and so on). Assume that the flood damage at location $i$ is a function of primarily the flood stage, $S$, at that location. Denote this potential damage function as $D_i(S)$. Such a function is illustrated in Figure D.31.

Integrating the product of the annual exceedance probability of flood stage, $F_s(S)$, and the potential damages, $D_i(S)$, over all stages $S$ will yield the annual expected damages, $E[D_i]$, for land parcel $i$;

$$E[D_i] = \int D_i(S)F_s(S)\mathrm{d}s \tag{D.11}$$

The sum of the expected damage estimates over all the parcels of land $i$ on the floodplain is the total expected damage that one can expect each year, on average, on the floodplain.

$$EAD = \sum_i E[D_i] \tag{D.12}$$

Alternatively, the annual expected flood damage could be based on a calculated probability of exceeding a specified flood damage, as shown in Figure D.22. For this method, the potential flood damages, $D_i(S)$, are determined for various stages $S$ and then summed over all land parcels $i$ for each of those stage values $S$ to obtain the total potential damage function, $D(S)$, for the entire floodplain, defined as a function of flood stage $S$:

$$D(S) = \sum_i D_i(S) \tag{D.13}$$

This is the function shown in quadrant (a) in Figure D.22.

Levee failure probabilities, $PLF'(p)$, based on the exceedance probability $p$ of peak flows, or stages, as defined in Figure D.29 can be included in calculations of expected annual damage. Expressing the damage function, $D(S)$, as a function, $D'(p)$, of the stage exceedance probability $p$ and multiplying this flood damage function $D'(p)$ times the probability of levee failure, $PLF'(p)$ defines the joint exceedance probability of expected annual damages. Integrating over all values of $p$ yields the expected annual flood damage, $EAD$:

$$EAD = \int_0^1 D'(p)PLF'(p)\mathrm{d}p \tag{D.14}$$

Note that the floodplain damage and probability of levee failure functions in Equation D.14 both increase with increasing flows or stages, but as peak flows or stages increase, their exceedance probabilities decrease. Hence, with increasing $p$ the damage and levee failure probability functions decrease. The effect of levees on the expected annual flood damage, $EAD$, is shown in Figure D.32. The 'without levee' function in the lower right quadrant of Figure D.28, is $D'(p)$. The 'with levee' function is the product of $D'(p)$ and $PLF'(p)$. If the probability of levee failure, $PLF'(p)$ function were as shown in Figure D.28, (that is, if it were 1.0 for values of $p$ below some overtopping stage associated with an exceedance probability $p^*$, and 0 for values of $p$ greater than $p^*$), then the function would appear as shown 'with levee, overtop only' in Figure D.32.

### 4.4.1. Risk-Based Analyses

Risk-based analyses attempt to identify the uncertainty associated with each of the inputs used to identify the appropriate capacity of various flood risk reduction measures. There are numerous sources of uncertainty associated with each of the functions shown in quadrants (a), (b) and (c) in Figure D.21. This uncertainty translates into uncertainty associated with estimates of flood risk probabilities and expected annual flood damage reductions obtained from reservoir flood storage capacities and channel improvements.

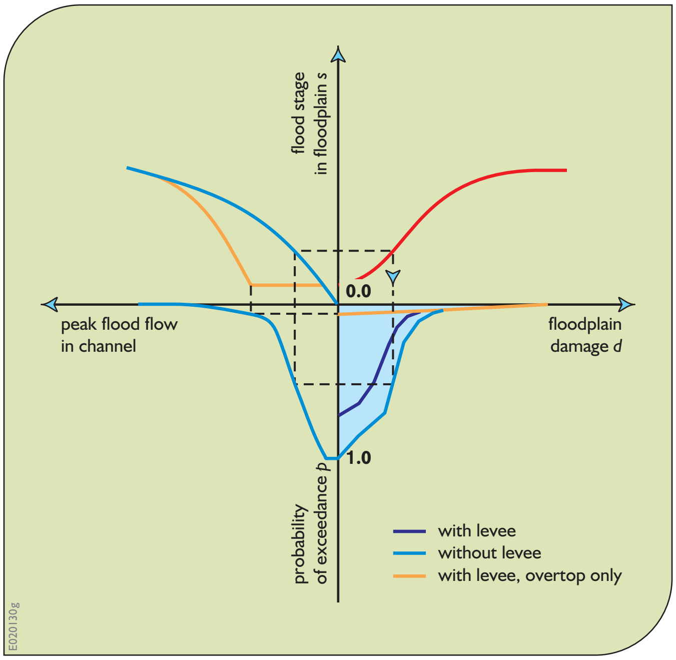



**Figure D.32.** Calculation of expected annual flood damage taking into account probability of levee failure.

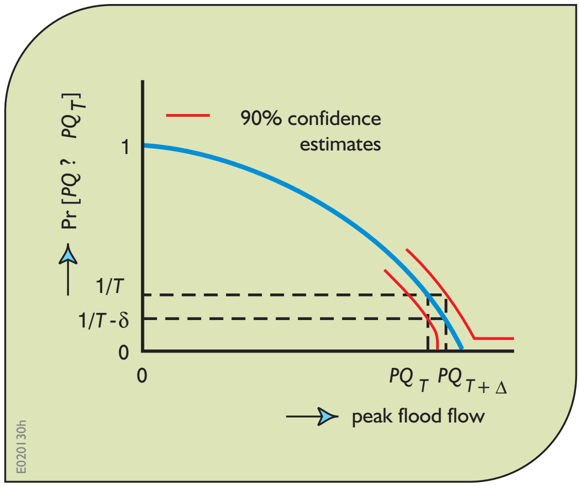



**Figure D.33.** Portion of peak flow probability of exceedance function showing contours containing 90% of the uncertainty associated with this distribution. To be 90% certain of protection from a peak flow of $PQ_T$, protection is needed from the higher peak flow, $PQ_{T+\Delta}$ expected once every $T+\Delta$ years, i.e. with an annual probability of $1/(T+\delta)$ or $(1/T)-\delta$ of being equalled or exceeded.

Going to the substantial effort and cost of quantifying these uncertainties, which themselves will be surely be uncertain, does provide additional information. The design of any flood protection plan can be adjusted to reflect the attitudes of stakeholders toward the uncertainty associated with specified flood peak return periods or equivalently their probabilities of occurring in any given year.

For example, assume a probability distribution capturing the uncertainty about the expected probability of exceedance of the peak flows at the potential damage site (as shown in Figure D.22) is defined from a risk-based analysis. Figure D.33 shows that exceedance function together with its 90% confidence bands near the higher flood peak return periods. To be, say, 90% sure that protection is provided for the $T$-year return period flow, $PQ_T$, one may have to design for an equivalent expected $T+\Delta$ year return period flow, $PQ_{T+\Delta}$; that is, the flow having a $(1/T)-\delta$ expected probability of being exceeded. Conversely, protection from the expected $T+\Delta$ year peak flood flow will provide 90% assurance of protection from flows that will occur less than once in $T$ years on average.

If society wanted to eliminate flood damage it could do so, but at a high cost. It would require either costly flood control structures or the elimination of economic activities on floodplains. Both reduce expected economic returns from the floodplain, and hence, are not likely to be undertaken. There will always be a risk of flood damage. Analyses such as those just presented help identify the risks. They can be reduced and managed but not eliminated. Finding the best levels of flood protection and flood risk, together with risk insurance or subsidies (illustrated in Figure D.34) is the challenge for public and private agencies alike. Flood risk management is as much concerned with good things not happening on floodplains as with bad things happening on them.

## 5. Decision Support and Prediction

To assist decision-makers in evaluating alternative solutions for problems encountered in integrated river management, and in communicating these solutions and their consequences to the public, interactive computer-based decision support systems (DSS) can be of considerable help. A DSS could consist of up-to-date river and floodplain information (using geographic information systems and database managers) and dynamic river

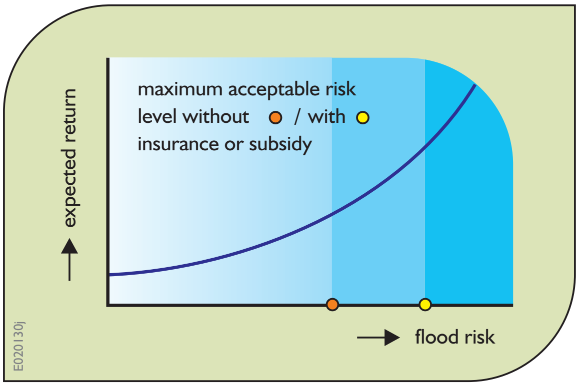



**Figure D.34.** Relationship between expected economic return from floodplain use and risk of flooding. The lowest flood risk does not always mean the best risk, and what risk is acceptable may depend on the amount of insurance or subsidy provided when flood damage occurs.

models of the complete river network. It could be used to simulate flooding scenarios in which water managers and stakeholders could be involved in deciding what to do and noting the impacts or consequences. Periodic simulation exercises such as this could serve as a reminder to the public, especially during dry years, that floods can and will occur, and some may have to be allowed to overflow levees and cause damage.

## 5.1. Floodplain Modelling

Efforts to identify improved flood management policies are almost always based on predictions, admittedly uncertain, of the impacts of specific structural or non-structural measures. These impact predictions come from models and their databases. Modelling and the communication of model results to interested property owners, government planners and insurance companies are becoming increasingly important components of flood management.

In providing the desired information, one must deal with uncertainties. Flood studies imply extrapolation of model parameters and boundary data to situations that have never been recorded. This raises the question of the reliability of current procedures of model calibration. The quality of model results depends to a large extent on available data and their relevance for future planning. Two-dimensional (2D) models require accurate digital terrain models (DTM). The cost of collecting data of sufficient accuracy for flood studies has decreased significantly through the use of airborne laser altimetry. Moreover, research has started to provide floodplain roughness data through complementary analysis of laser altimetry data. Laser altimetry techniques are now also applied to the recording of flood levels, which can further support the calibration of flood simulation models. The new data collection techniques are also complemented by hydraulic laboratory investigations, such as the derivation of laws and parameters describing flow resistance resulting from various classes of vegetation.

New data collection techniques also show a distinct change in the type of data underlying model development and model use. In general, there is a tendency to move from the collection of point data to spatially-distributed data. Whereas design flood hydrographs used to form the basis for flood control studies, the rainfall radar data, together with river catchment models, allow the use of design storm patterns. The development of 2D models based upon digital terrain models and roughness maps is becoming easier and more cost-effective than the development of 1D models based upon point information of cross sections measured at selected intervals along the river axis.

This does not mean that the traditional 1D modelling of flood events will be fully replaced by 2D model applications. Although the development of an accurate 2D model supported by the new data collection and processing techniques is becoming easier and cheaper for many studies, 2D models require more computer time, and this is not going to change within a decade. Flood management planning and risk assessments typically require the simulation of many floodplain management alternatives over large numbers of probabilistic or stochastic events.

Both 1D and 2D models can be used together as appropriate. They can be integrated in many applications to provide local refinement of the description, in particular the flooding of normally dry land. A 2D model can be used as a support tool for the development of accurate and extrapolatable 1D models through the transfer of discharge conveyance and storage functions.

The shift from a policy of protecting as much agricultural, industrial and urban land as possible by constructing dykes or levees close to the main river, to one of creating, if possible, green rivers and larger floodplains or areas with lower protection against flooding results in a need to understand and predict the flood flows and heights resulting from such measures. This requires data on the local topography and flow resistance.

The hydraulic roughness of floodplains can be assessed with the aid of maps showing vegetation or land use classes. Such habitat or land use compositions can be related to flow roughness descriptions. Assessments of roughness values associated with each vegetation class are usually based upon a combination of hydraulic laboratory experiments and expert judgement. The spatial distribution of vegetation or land use classes subsequently leads to the assignment of roughness values at the grid cells of 2D models. Flow resistance for each vegetation class can be constant or expressed as a function of stage (flood height).

The 2D models incorporating these floodplain roughness parameters can be used to assist in the calibration of 1D models. Both types of model can then be used for a variety of water management studies and policy analyses.

The link between vegetation or land use classes and roughness provides a description of the energy losses as a function of the flow-blocking area in a specific region. This provides a more accurate description of the losses as a function of the water level. Flow through reed fields, for example, meets a high resistance as long as the vegetation is not overtopped or bent-over and lying on the bottom. Other, more dense vegetation may even create laminar flow before overtopping takes place. After drowning, its flow resistance reduces significantly. A relation between vegetation or land use and flow resistance also allows for the introduction of seasonal variations in flow resistance.

One software product developed at WL | Delft Hydraulics is the so-called Delft-1D2D system. This is currently in use for the simulation of large numbers of scenarios of flooding resulting from potential dyke breaches. The approach followed in this software provides many possibilities for the integrated simulation of river floods, tidal flow and urban storms.

## 5.2. Integrated 1D–2D Modelling

There are various ways in which one can couple hydrological and hydraulic system simulation models. These include:

- Sequential coupling of data flows if one process is completely independent of another process, for example, from a rainfall–runoff module to a channel flow module.
- On-line explicit coupling at a selected frequency in time, such as the link between a surface water flow module and a water quality module. Such coupling works well if the mutual dependence of processes described by the individual modules is not important: for example, if the time scale of one process is quite different from the time scale of the other process. If process time scales are similar, constraints in time stepping may have to be satisfied.
- On-line implicit coupling of modules. This approach requires not only the exchange of data between various modules but also the exchange of complete sets of equations, which are to be solved simultaneously.

The WL | Delft Hydraulics' general 0D, 1D and partly 2D modelling system called SOBEK is designed to integrate processes, such as rainfall–runoff; river, channel and pipe flow; water quality and morphology. The hydrodynamic part handles sub- and supercritical flows, including transitions such as hydraulic jumps and tidal bores.

This approach to integration offers a range of modelling options. In an urban drainage model, one can combine pipe flow, river flow, tidal flow, street flow, and flow over parks and parking places, for example (Bishop and Catalano, 2001). Another example is dam break analysis that allows for any suitable combination of 1D and 2D flow, including zones with supercritical flow. Finally, the creation of retention storage in flood control is modelled more realistically than on the basis of a combined 0D and 1D schematization by itself.

The 1D2D integration can be illustrated by examining alternative flood scenarios for the Geldersche Vallei in the centre of the Netherlands (Figure D.35). The area modelled measures approximately 35 × 20 km, schematized with a 2D grid of 100 × 100 m and coupled to a 1D schematization of the principal waterways and road culverts in the area. The model study had the objective of assessing flood damage and establishing evacuation plans for a number of flood scenarios. The most severe situation is caused by a possible breach in the Grebbedijk, protecting the area from floods conveyed along the Lower Rhine.

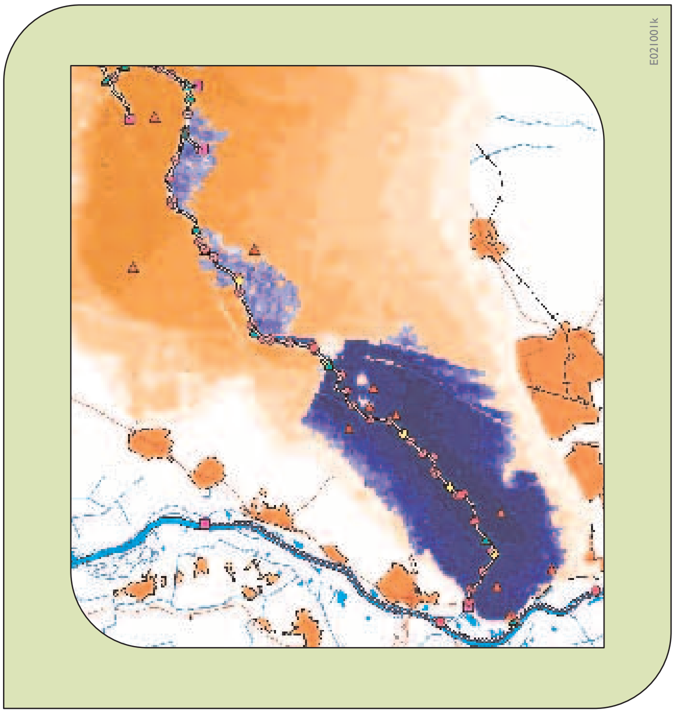



**Figure D.35.** Coupled 1D2D SOBEK model of the Geldersche Valley in the Netherlands.

The coupled model also allows for the extension of the 2D model along the Lower Rhine. This river has been schematized in 1D over a length of 19 km, with the breach coupled a few kilometres downstream of the upstream boundary of the Lower Rhine branch. The flow through the breach significantly influences water levels upstream, and consequently the discharges generated through it. It is here that another advantage of implicit coupling shows up. Inclusion of the Lower Rhine into the 2D model would have given only approximately a 3 to 10 grid-point resolution over its width, which is not sufficient to model its flow satisfactorily. Another and often principal advantage of 1D and 2D coupling is that the grid cells of the 2D model part can generally be significantly larger than when pure 2D modelling is applied. Apart from the flexibility in schematization, this usually also leads to considerable savings in simulation times.

# 6. Conclusions

In the wake of the unusually severe floods on the River Rhine and on the upper Mississippi River in the early and mid-1990s and the equally severe floods in many parts of the world since then, including those in Central Europe and in Asia that occurred in 2002 and 2004, an increasing consensus is emerging. Humans must provide space for floods where and when they occur, and floods must be managed better. This is not a new idea. The only sustainable way to reduce the continuing cycle of increasing flood damage and increasing flood protection costs is to restore the natural floodplains that contain the river's floodwaters. Floods have become a reason to restore rivers rather than dam or dyke them. Many who have traditionally advocated structural solutions to flood protection are increasingly embracing the idea that rivers need to be given space to flood. Flood protection has joined clean drinking water and recreation as a reason for restoring river ecosystems.

Governments can motivate more responsible floodplain management. National governments can assume leadership, define clear roles and provide fiscal support for flood management at lower levels. As happened in the upper Mississippi River Basin, governments can

- provide relocation aid and buyout funding for those living in floodplains
- stop disaster relief payments and subsidized flood insurance for those who continue to live or develop in flood prone areas
- make flood insurance mandatory where offered, and have the premiums reflect flood risks
- enforce building code requirements on floodplains.

In the United States, the Army Corps of Engineers Flood Control Programme has had a major influence on floodplain development. Many now might claim it to be a negative influence, but the Corps did what Congress, and the public, asked them to do. Hundreds of dams and thousands of miles of Corps levees and floodwalls were built for flood protection. The result, of course, was to encourage further development in flood-prone areas. Existing Corps projects continue to influence the management of most major rivers and their floodplains, including those of the Mississippi, Missouri, Ohio and Columbia River Basins. Although local government is ultimately responsible for decisions regarding land use in the United States, flood control projects constructed by the Corps provide incentives for floodplain development.

The Corps' analyses of benefits and costs have traditionally strongly favoured structural flood control projects. Engineers like to build things. Many within the Corps now recognize that their regulations and incentives need to be changed to enable them to take a more balanced view and allow non-structural flood control projects to compete with structural ones.

In the Netherlands, a similar story can be told. There is an expression the Dutch like to say: 'God made the Earth, but the Dutch made the Netherlands.' Without structures, much of the country would be permanently under water. But here, as in the United States and in most of the rest of the world where floods occur, flood management needs to become a part of integrated water and land management. Multiple uses, multiple sectors of the economy, and the interests of multiple stakeholders need to be considered when developing comprehensive land use and water management plans and policies.

Options available to local governments for floodplain management include land use zoning. Land zoning should be based on comprehensive land use plans that define a vision of how a community should be developed (and where development should not occur). Through these plans, uses of the land can be tailored to match the land's economic benefits as well as its hazards. For example, flood hazard areas can be reserved for parks, golf courses, backyards, wildlife refuges, natural areas or similar uses that are compatible with the natural flooding process.

Open space preservation should not be limited to floodplains, because some sites in the watershed (but outside the floodplain) may be crucial to controlling runoff that adds to the flood problem. Areas that need to be preserved in a natural state should be listed in land use and capital improvement plans.

Zoning and open space preservation are ways to keep damage-prone development out of hazardous or sensitive areas. Floodplain development regulations can include construction standards governing what can and cannot be built in the floodplain. They can serve to help protect buildings, roads and other projects from flood damage, and prevent development from aggravating the flood problem. The three most common types of floodplain regulations are subdivision ordinances, building codes, and 'stand-alone' floodplain ordinances.

Several measures can help reduce runoff of stormwater and snowmelt throughout the watershed. Retention and detention regulations, usually part of a subdivision ordinance, require developers to build retention or detention basins to minimize the increases in runoff caused by new impervious surfaces and new drainage systems. Best management practices (BMPs) reduce polluted runoff entering waterways. Pollutants in runoff may include lawn fertilizers, pesticides, farm chemicals, and oils from street surfaces and industrial areas.

Wetlands filter runoff and adjacent surface waters to protect the quality of lakes, bays and rivers, and protect many of our sources of drinking water. They can store large amounts of floodwater, slowing and reducing downstream flows. They can protect shorelines from erosion. They also serve as a source of many commercially and recreationally valuable species of fish, shellfish and wildlife.

Moving a flood-prone building to higher ground is the surest and safest way to reduce its risk from flooding. Acquisition of flood-prone property is undertaken by a government agency, so the cost is not borne by the property owner. After any structures are removed, the land is usually converted to public use, such as a park, or allowed to revert to natural conditions. There are a variety of funding programmes that can support a local acquisition project. For example, more than 8,000 homes were acquired or relocated by the US Government after the 1993 Mississippi Flood.

Based on lessons learned from floods and flood protection efforts throughout the world, the following principles are suggested:

- Restore river systems and functions that improve flood management, while at the same time restoring the natural waterway and its ecosystems:
  - Restore to a meaningful extent the historic capacity of rivers and their floodplains to better accommodate floodwaters by setting back levees to widen the floodway of the river channel.
  - Increase wetland and riverside forest habitat within the widened river zone.
  - Increase the use of planned floodplain flooding to reduce downstream flood peaks.

  - Strengthen existing and properly sited levees at high risk that protect high value floodplain uses that cannot be relocated from the floodplain.
  - Re-assess the operations of reservoirs and waterworks to ensure efficient, reliable and prudent use of flood control space. In some cases, dams and waterworks need to be structurally modified to improve their ability to release water to avoid downstream flooding.
  - Improve use of weather forecasting and monitoring upstream conditions so as to have a better 'early warning system' of when a flood could be coming.
- Manage the uses of floodplains to minimize taxpayer expense and maximize environmental health:
  - Eliminate incentives or subsidies for development in the most dangerous parts of the floodplain. No more people should be put in harm's way.
  - Reform floodplain mapping programmes so that they accurately portray the risks and consequences of anticipated flooding. Ensure that people understand the risk of flooding where they are located.
  - Ensure that new structures that have to be built in floodplains are designed to resist damage from foreseeable future floods.
  - Educate people in the risks of living, working or farming in areas prone to floods, and make sure they are willing to bear the appropriate financial responsibility for such use.
  - Endeavour to relocate the most threatened people and communities who volunteer to move to safer locations.
  - Ensure that state and local governments responsible for floodplain land use decisions bear an increased financial responsibility for flood recovery efforts.
- Manage the entire watershed to provide the most protection from floods in an environmentally sensitive way:
  - Discourage development in remaining wetlands and floodplains. Wetlands and functioning floodplains act as giant sponges to absorb and slow the progress of floodwaters.
  - Use acquisition and easement programmes to restore historical wetlands and floodplain acreage and to promote the functional restoration of associated river systems.
  - Discourage clearcutting and road building in areas prone to landslides.
  - Where possible, replace non-native hillside annual vegetation with native perennials to improve rainwater absorption and reduce hillside erosion.

Newly developed technologies are facilitating the development of models for studying the propagation of floods through floodplains, coastal zones and urban areas. They include the use of airborne laser altimetry to provide cheaper and more accurate digital terrain models. In addition, the technology is being used for an efficient assignment of flow resistance parameters to flood models. It also offers further support to model calibration by monitoring water levels during the passage of flood waves.

The descriptive capabilities of flood models have been extended significantly through the implicit coupling of 1D and 2D schematizations. Integrated 1D–2D surface and subsurface modelling can be applied to floodplain and urban flooding investigations. These applications can lead to a better understanding of the flood phenomena, more accurate predictions and better planning.

## 7. References

BISHOP, W.A. and CATALANO, C.L. 2001. Benefits of two-dimensional modelling of urban flood projects. In: *Sixth International Conference on Hydraulics in Civil Engineering*, 28–30 November 2001, Hobart, Australia. Barton, Australia, Institution of Engineers.

INTERAGENCY FLOODPLAIN MANAGEMENT REVIEW COMMITTEE (Galloway Committee). 1994. *Sharing the challenge: floodplain management into the twenty-first century*. Washington, D.C., US Government Printing Office.

USACE. 1991. *Proceedings of a Hydrology and Hydraulics Workshop on Riverine Levee Freeboard*, 27–29 August 1991, Monticello, Minnesota. Davis, Calif., USACE Hydrological Engineering Center.

WHITE, G. 1945. *Human adjustments to floods*. Chicago, Ill., University of Chicago Press.

WLOSINSKI, J.; OLSEN, D.A.; LOWENBERG, C.; OWENS, T.W.; ROGALA, J. and LAUSTRUP, M. 2002. Habitat changes in the Upper Mississippi River floodplain. National Biological Service, http://biology.usgs.gov/s+t/noframe/m2136.htm (accessed 16 November 2004).

## Additional References (Further Reading)

ALVAREZ, F. and VERWEY, A. 2000. Simulation of wind set-up on lakes in a one-dimensional model schematisation. *Fourth International Conference on Hydroinformatics*. 23–27 July 2000, Cedar Rapids, Iowa. Iowa City, Iowa Institute for Hydraulic Research (CD-rom).

ASSELMAN, N.E.M. 2001. Laser altimetry and hydraulic roughness of vegetation. Arnhem, the Netherlands, Report Q2701 prepared for RIZA of the Directorate-General for Public Works and Water Management.

ASSELMAN, N.E.M.; MIDDELKOOP, H.; RITZEN, M.R.; STRAATSMA, M.W.; VAN VELZEN, E.H. and JESSE, P. 2002. Assessment of the hydraulic roughness of river floodplains using laser altimetry. In: F. Dyer, M.C. Thoms and J.M. Olley (eds). *The Structure, Function and Management Implications of Fluvial Sedimentary Systems*. IAHS International Symposium on the Structure, Function and Management Implications of Fluvial Sedimentary Systems, 2–6 September 2002, Alice Springs, Australia. International Association of Hydrological Sciences (Publication no. 276).

CARLSON, E. 1996. Turn environmental confrontation into cooperation. *LandOwner Newsletter*, 9 Dec.

COLLINS, N. and CATALANO, C.L. 2001. Specialised 2D modelling in floodplains with steep gradients. *Sixth International Conference on Hydraulics in Civil Engineering*, 28–30 November 2001, Hobart, Australia. Barton, Australia, Institution of Engineers.

DUEL, H.; PEDROLI, B. and LAANE, W.E.M. 1996. The habitat evaluation procedure in the policy analysis of inland waters in the Netherlands: towards ecological rehabilitation. In: M. Leclerc, H. Capra, S. Valentin, A. Boudreault and Y. Côté (eds). *Ecohydraulics 2000*, Proceedings of the Second IAHR Symposium on Habitat Hydraulics, Québec (INRS-Eau, Sainte-Foy, Qc), June 1996. Québec, INRS-Eau. pp. A619–30.

DUFF, I.S.; ERISMAN, A.M. and REID, J.K. 1986. and *Direct methods for sparse matrices*. Oxford, UK, Claredon.

FAPRI (Food and Agricultural Policy Research Institute). 1994. *The economic impact of a loss of navigation on the Missouri River on Missouri agriculture*. Ames, Iowa, Iowa State University.

FRANK, E.; OSTAN, A.; COCCATO, M. and STELLING, G.S. 2001. Use of an integrated one dimensional–two dimensional hydraulic modelling approach for flood hazard and risk mapping. In: R.A. Falconer and W.R. Blain (eds). *River basin management*, Southampton, UK, WIT. pp. 99–108.

GOLUB, G.H. and VAN LOAN, C.F. 1983. *Matrix computations*. Oxford, UK, North Oxford Academic.

GOMES PEREIRA, L.M. and WICHERSON, R.J. 1999. Suitability of laser data for deriving geographical information: a case study in the context of management of fluvial zones. *ISPRS Journal of Photogrammetry and Remote Sensing*, No. 54, pp. 105–14.

HARDIN, G. 1968. The tragedy of the commons. *Science*, No. 162, pp. 1243–48.

SAST (Scientific Assessment and Strategy Team). 1994. *A blueprint for change, part v: science for floodplain management into the twenty-first century*. Washington, DC, SAST.

SILVA, W.; KLIJN, F. and DIJKMAN, J.P.M. 2001. *Room for the Rhine Branches in the Netherlands*. Arnhem and Delft, the Netherlands, Directorate-General for Public Works and Water Management and WL | Delft Hydraulics.

STELLING, G.S.; KERNKAMP, H.W.J. and LAGUZZI, M.M. 1998. Delft flooding system: a powerful tool for inundation assessment based upon a positive flow simulation. In: V. Babovic and L.C. Larsen (eds.), *Hydroinformatics '98*, Rotterdam, Balkema. pp. 449–56.

USACE. 1992. *Final Upper Mississippi River navigation study: reconnaissance report, revised version*. Rock Island, Ill., USACE.

USACE. 2001. *Missouri River master water control manual review and update: draft environmental impact statement, executive summary*. Omaha, Neb.

VERWEY, A. 1994. Linkage of physical and numerical aspects of models applied in environmental studies. Keynote lecture. In: *Proceedings of the Conference on Hydraulics in Civil Engineering*. International Conference on Hydraulics in Civil Engineering, 15–17 February 1994, Brisbane, Australia. Australian Institution of Engineers (National Conference Publication 94/1).

VERWEY, A. 2001. Latest developments in floodplain modelling: 1D/2D integration. In: *Proceedings of the Sixth Conference on Hydraulics in Civil Engineering, 28–30 November 2001, Hobart, Australia*. Hobart, Institution of Engineers.

WELTZ, M.A.; RITCHIE, J.C. and FOX, H.D. 1994. Comparison of laser and field measurements of vegetation height and canopy cover. *Water Resources Research*, No. 30, pp. 1311–19.

WHITE, G. 1986. Geography, resources, and environment. In: W. Kates and I. Burton (eds). *Volume I: Selected writings of Gilbert White*, Chicago, Ill.: University of Chicago Press. pp. 443–59.

## Appendix E: Project Planning and Analysis: Putting it All Together


1. Basic Concepts and Definitions 645
   1.1. The Water Resources System 645
   1.2. Functions of the Water Resources System 646
      1.2.1. Subsistence Functions 646
      1.2.2. Commercial Functions 646
      1.2.3. Environmental Functions 647
      1.2.4. Ecological Values 647
   1.3. Policies, Strategies, Measures and Scenarios 647
   1.4. Systems Analysis 648

2. Analytical Description of WRS 649
   2.1. System Characteristics of the Natural Resources System 650
      2.1.1. System Boundaries 650
      2.1.2. Physical, Chemical and Biological Characteristics 650
      2.1.3. Control Variables: Possible Measures 651
   2.2. System Characteristics of the Socio-Economic System 651
      2.2.1. System Boundaries 651
      2.2.2. System Elements and System Parameters 651
      2.2.3. Control Variables: Possible Measures 652
   2.3. System Characteristics of the Administrative and Institutional System 652
      2.3.1. System Elements 652
      2.3.2. Control Variables: Possible Measures 652

3. Analytical Framework: Phases of Analysis 652

4. Inception Phase 654
   4.1. Initial Analysis 655
      4.1.1. Inventory of Characteristics, Developments and Policies 655
      4.1.2. Problem Analysis 655
      4.1.3. Objectives and Criteria 656
      4.1.4. Data Availability 656
   4.2. Specification of the Approach 657
      4.2.1. Analysis Steps 657
      4.2.2. Delineation of System 657
      4.2.3. Computational Framework 658
      4.2.4. Analysis Conditions 659
      4.2.5. Work Plan 660
   4.3. Inception Report 660
   4.4. Communication with Decision-Makers and Stakeholders 661

5. Development Phase 661
   5.1. Model Development and Data Collection 661
      5.1.1. Analysis of the Natural Resources System (NRS) 661
      5.1.2. Analysis of the Socio-Economic System (SES) 664
      5.1.3. Analysis of the Administrative and Institutional System (AIS) 666
      5.1.4. Integration into a Computational Framework 667
   5.2. Preliminary Analysis 668
      5.2.1. Base Case Analysis 669
      5.2.2. Bottleneck (Reference Case) Analysis 669
      5.2.3. Identification and Screening of Measures 669
      5.2.4. Finalization of the Computational Framework 669

6. Selection Phase 670
   6.1. Strategy Design and Impact Assessment 670
   6.2. Evaluation of Alternative Strategies 671
   6.3. Scenario and Sensitivity Analysis 672
   6.4. Presentation of Results 672

7. Conclusions 672

# Appendix E: Project Planning and Analysis: Putting it All Together

**The main purpose of water resources management is to ensure the best use of available water resources. In addition to supporting life itself, water is used in the production of economic goods and services that are needed to meet national and regional development goals. Planning projects are often needed to determine how best to develop and manage these resources. This appendix describes the general approach used by WL | Delft Hydraulics to assess water resources systems and develop management strategies for them. Each water resources system is different and has different problems, and the specific application of any planning approach should address the particular issues involved. What is important in all cases is the comprehensive and systematic process of analysis, together with constant communication among planners, decision-makers and the interested and affected public.**

## 1. Basic Concepts and Definitions

The approach taken for water resources planning involves a number of terms and concepts related to planning, management and the role of systems analysis. These terms and concepts are commonly used in water resources planning, but with varying connotations. To avoid confusion, we begin this appendix by indicating what we mean when using these terms and concepts.

### 1.1. The Water Resources System

A water resources system (WRS) can be considered to consist of:

- *The natural resources system* (NRS):
  - the natural sub-system of streams, rivers, lakes and their embankments and bottoms, and the groundwater aquifer
  - the infrastructure sub-system, such as canals, reservoirs, dams, weirs, sluices, wells, pumping plants and wastewater treatment plants (including the operation rules for elements of this sub-system)
  - the water itself, including its physical, chemical and biological components in and above the soil, often referred to as the 'ABC' components: abiotic or physical, biological, and chemical.
- *The socio-economic system* (SES):
  - water-using and water-related human activities.
- *The administrative and institutional system* (AIS):
  - the system of administration, legislation and regulation, including the authorities responsible for managing and implementing laws and regulations.

This definition of a WRS covers the aspects that are essential for natural resources management: supply, demand and means to manage the resources. The NRS incorporates the supply side of the system (resource base), and the SES the demand side. The management of both the supply and the demand sides is provided by the AIS.

The external natural boundaries of an NRS usually consist of the water divides of the catchment area, boundaries of the groundwater aquifer(s) belonging to it, and the point where the river or canal discharges into the sea. Beyond the latter point, one usually speaks of coastal waters. There is a (brackish) transition zone between the two systems. When discharges are large, the brackish water zone moves seaward. The extent to which the brackish water zone is explicitly included in the NRS depends upon the interests involved, the problems encountered and the management objectives to be attained.

The geographical boundaries of the SES and the AIS vary depending on what part of the socio-economic system is considered essential for managing and assessing the impacts of the NRS. For example, in order to analyse the relations between an NRS and a region with an open economy, one would have to examine the relationship of this region with the economy outside it so as to estimate future developments in the region.

## 1.2. Functions of the Water Resources System

A general framework of functions is presented in Table E.1. The classification distinguishes between the functions and uses that are tangible and those that are not. Tangible functions, such as hydropower generation or municipal water supply, may be assigned a monetary value; intangible functions are activities such as nature conservation. In between are environmental functions, some of which may be given a direct value and others valued indirectly, by using a shadow price or other valuation method (see Chapter 10). The self-purification process of a river, for example, may be assigned a shadow price by comparing this 'work done by nature' with the costs of constructing and operating a wastewater collection and treatment system. The functions in Table E.1 are explained below.

### 1.2.1. Subsistence Functions

Local communities depend to a large extent on ambient water resources for household uses, and for irrigating home gardens and village irrigation plots. They may also use the streams, paddy fields, ponds and lakes for fishing. These uses are often neglected in national economic accounts, as they are not marketed or otherwise assigned a monetary value. However, if the WRS becomes unable to provide these products, this may well be considered an economic loss, as the people who are dependent on these products now have to buy them. An example is the cost of providing purified drinking water where water quality has deteriorated and become undrinkable.

### 1.2.2. Commercial Functions

Commercial uses of water resources are reflected in national economic accounts because they are marketed or otherwise given a monetary value, e.g. the price to be paid for domestic water supplies. Catching fish for sale by individuals and commercial enterprises is an example. These uses have a commercial value and most are also consumptive in nature.

The concept of 'non-consumptive use' should be regarded with certain reservations. Non-consumptive water use may alter the performance of the WRS in way

**Table E.1.** Functions of the water resources system.

| Function | Description | Examples |
|---|---|---|
| Subsistence functions | Local communities make use of water and water-based products which are not marketed | – local drinking water supply<br>– traditional fishing<br>– subsistence irrigation |
| Commercial functions | Public or private enterprises make use of water or water-based products that are marketed or otherwise given a monetary value | – urban drinking water supply<br>– industrial water supply<br>– irrigation<br>– hydro-power generation<br>– commercial fishing<br>– transportation |
| Environmental functions | Regulation functions<br>Non-consumptive use | – purification capacity<br>– prevention of salt intrusion<br>– recreation and tourism |
| Ecological values | Values of the WRS as an ecosystem | – integrity<br>– gene pool, biodiversity<br>– nature conservation value |

that constrains or increases the cost to other users. Hydroelectric generation is an example of a partly non-consumptive use that affects the system in various ways. First, reservoirs built for hydropower increase evaporation losses, and hence reduce the amount of water available for downstream users. Second, operation of the reservoir for the production of 'peak power' may alter the flow regime downstream, and this can adversely affect downstream habitats and users. Finally, water quality problems related to reservoirs may seriously affect the ecosystems both upstream and downstream of the reservoir.

Another example of partly non-consumptive use is inland water transportation. Oil and chemical pollution caused by water transport activities can affect other users and the ecosystem that depend upon the water resources. Moreover, inland water transportation may involve a real consumptive demand for water. If water depths are to be maintained at a certain level for navigational purposes, releases from reservoirs may be required which provide no value to other water users. An example is the Lower Nile system, where water is released from Lake Nasser to enable navigation (and energy generation) during the so-called winter closure. This water could otherwise remain stored for (consumptive) use by agriculture during the growing season.

#### 1.2.3. Environmental Functions

The drainage basin of a river fulfils a series of functions that require no human intervention, and thus have no need of regulatory systems. These environmental functions include the self-purification capacity of the water system, and recreational and tourism uses. It is sometimes difficult to assign a value to environmental functions. They may be assessed by using a shadow price, calculated as the costs of providing similar functions in other ways, e.g. the cost of additional wastewater treatment. Recreational and tourism values may be determined by assessing the economic benefits accruing from the use of tourist facilities like hotels, and/or the revenues from fishing licenses.

#### 1.2.4. Ecological Values

Water is a substance that is essential for life. Rivers, streams and lakes not only offer an environment for aquatic species, but are often bordered by wetlands such as reed beds, floodplains and marshes. These land–water ecotones (transition areas between adjacent ecological communities) are known to harbour a rich assemblage of species, and are also important for the diversity of adjacent ecological communities. These ecological entities have a value of their own, irrespective of the actual or potential human use, an intrinsic ecological value. There are many concepts and expressions that describe this ecological value: 'heritage value', 'aesthetic value', 'nature value', 'option value', 'existence value' among others.

Environmental functions relate to the benefits of the natural environment for humans, and ecological values include the intrinsic value of nature. In view of the increasing emphasis on sustainability, this distinction is useful in water resources planning and management. It points clearly to the need for the continuous care for the natural conditions of our planet and the maintenance of an acceptable and sustainable level of environmental quality.

### 1.3. Policies, Strategies, Measures and Scenarios

In planning the terms *policy goal*, *strategy*, *measure* and *scenario* are frequently used. In popular use they are often treated as interchangeable, and this can be confusing. In this appendix, and indeed in this book, the following meanings are used:

- A *policy goal* has to do with identifying the needs, prioritizing issues and setting targets for sectors or regions. A policy goal does not by itself lay down specific actions; it merely sets the targets and constraints for the actions (levels, timing and budget). Policies also may specify in general terms how a certain target should be achieved: for example, by applying user-oriented demand management measures rather than relying on large-scale water supply infrastructure development. In specifying how targets should be achieved, other policies are taken into account (to do with foreign trade perhaps, or fiscal or income tax requirements, or environmental controls). Government policies are usually considered as given

**Box E.1. Definitions**

**Policy goal:** where do we want to go
**Strategy:** how do we want to get there
**Measure:** what are we going to do
**Scenario:** external development, affecting our strategy

for any WRS study, but the results of the study may lead to the adoption of new government policies.

– A *strategy* is defined as a logical combination of individual measures or decisions that provides a solution to the WRS problem, for example, the construction of a reservoir plus the widening of the canal downstream and the increase of the intakes of the irrigation system. Together, these measures will reduce the risk of damage to the agriculture sector in a drought-stricken area. An alternative strategy might be to implement a cropping pattern that uses less water.

• A *measure* is an individual management action or decision. A distinction can be made between:
  – *Technical (or structural) measures:* modifications of elements of the water resources infrastructure such as canals, pumping stations, reservoirs, and fish stairs or ladders. Technical measures often include managerial measures as described below.
  – *Managerial measures* to improve the (daily) operation of the system, such as better ways of using the infrastructure (reservoirs, gates, weirs, sluices, etc.).
  – *Ecological (non-structural) measures* to improve the functioning of the ecosystem, for example by introducing fry in spawning areas, or large herbivores.
  – *Economic incentives* to induce water consumers to use the water resources in a socially desired manner by changing the price of the resource use (through charges, taxes or subsidies).
  – *Regulation measures* to restrict uncontrolled use of the water resources (through land-use zoning, permits, pollution control and other forms of restrictive legislation).
  – *Institutional arrangements* specifying which governmental agencies are responsible for which functions of the WRS, and specifying the necessary interactions between public and private sectors involved.

• A *scenario* is defined as a development exogenous to the water system under consideration: in other words, developments that cannot be controlled by the decision-makers involved in the system. Examples of scenario variables are climate, climate change, demographical trends and changes, and economic growth. What should be treated as a scenario and what as a (potential) measure may depend on the system boundaries that have been set. In 'real' integrated water resources management studies (see Chapter 1, Section 5.1.3), restrictions on demographic and economic developments could be treated as potential actions. In that case they are not scenario variables anymore but should be treated as measures.

## 1.4. Systems Analysis

In relation to the systems approach to planning, the literature often emphasizes the mathematical techniques at the core of this approach. The use of mathematical tools, however, does not of itself constitute a systems approach. The approach, designed especially for complex systems of many interdependent components, involves:

• building predictive models to explain system behaviour
• devising courses of action (strategies) that combine observations with the use of models and informed judgments
• comparing the alternative courses of action available to decision-makers
• communicating the results to the decision-makers in meaningful ways
• recommending and making decisions based on the information provided
• monitoring and evaluating the results of the strategies implemented.

Systems analysis and policy analysis are often considered to be one and the same. If a distinction should be made, one might define systems analysis as an activity that does not apply only to policy problems. It can be applied to any system one wants to analyse for whatever reason.

System diagrams or conceptual models are important tools in systems analysis. A system diagram represents cause–effect relations between elements or sub-systems of the overall system. An example of the use of system diagrams in analysing water resources problems is presented in Figure E.1.

As the figure shows, water-using activities may cause two problems. First, the quantity demanded is greater than the supply; second, they affect the natural system (e.g. generate pollution or alter the water level) with undesired effects. The perception of these problems can be a trigger for analysis and planning activities, which in turn can result in management actions. The figure shows that the problem can be addressed in two ways: either by implementing demand-oriented measures (addressing the water use, i.e. SES), or by developing the water resources (i.e. NRS). Demand-oriented measures aim to reduce water use and effluent discharge per unit of output. Supply-oriented measures on the other hand

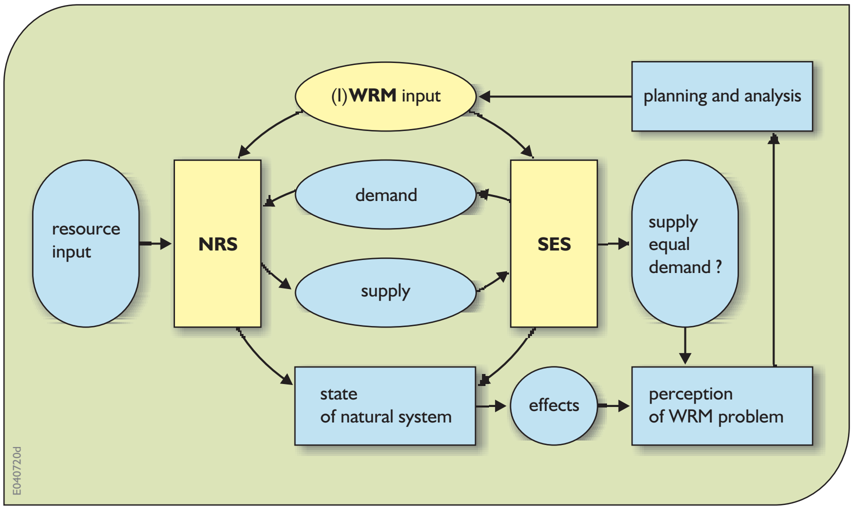



**Figure E.1.** Identification of a water resources management (WRM) problem.

are aimed at increasing the water supply so that shortages are reduced or at increasing the assimilative capacity of the receiving water bodies. Which measure or combination of measures is most effective depends on the criteria used by the implementing authority.

The relations between the elements of the systems can be expressed quantitatively. The system diagram is an outline of the computational framework to be used to analyse the performance of the system.

## 2. Analytical Description of WRS

Water resources management aims to match the usage of the WRS (the socio-economic system or SES) with the natural resources system (the NRS). This matching is based upon a common notion that scarce resources should be used for society in an 'optimal' way. Just what is optimum is determined by those who make decisions. The interaction between the use and availability of resources is controlled by the administrative and institutional system (the AIS) through which this management is implemented. These three 'entities' are depicted in Figure E.2.

The AIS 'manages' the other two entities, using information about the present or expected future state of each system. The management actions of this system are depicted by the single arrows leading towards the two interacting systems: by supply-oriented measures (infrastructure related) for the NRS and demand-oriented measures for the SES. The single arrow is not meant to imply there is no feedback from the NRS and the SES to the AIS; there must be feedback in the form of information

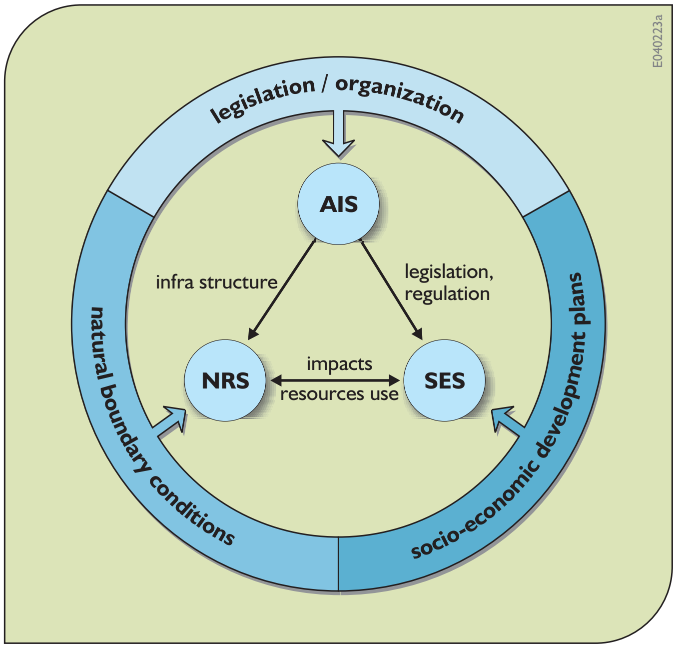



**Figure E.2.** Context for water resources planning.

on the state/performance of the two systems, otherwise effective management would be impossible. The arrows in Figure E.2 represent only the actions, not the information flows. The interaction between the user system (SES) and the resources system (NRS) is a physical one, depicted by the double arrow between the two systems. Resources provided by the NRS are used to support the SES, and the SES has physical effects on the NRS.

Each of the three systems is embedded within its own environment. The NRS is bounded by climate and (geo)physical conditions; the SES is formed by the demographic, social and economic conditions of the surrounding economies; and the AIS is formed and bounded by the constitutional, legal and political system. It is important to analyse these boundary conditions. Boundary conditions are usually considered fixed, but may not be as rigid as sometimes thought. For example, climatic conditions are mostly taken as fixed, and the analysis is based on historical information about precipitation, temperature and so on. These conditions may be considered to be subject to change, due to global warming. Whether to consider this possibility or not should be decided at the start of any analysis of a water management problem.

Economic growth is another example of a boundary condition that is often treated as fixed. If the resources system cannot sustain these foreseen developments (or only at very high costs), it may be appropriate to reconsider this fixed condition. By showing the consequences of unrestricted growth, the planning agencies can question the desirability of such development at a higher (usually national) planning level. This is represented in Figure E.2 by the border frame 'socio-economic development plans'. In fact, the arrow pointing inwards to the socio-economic system is reversed in such a case: the analysis provides information to a higher planning level that determines the (original) boundary conditions.

## 2.1. System Characteristics of the Natural Resources System

The natural resources system (NRS) is defined by its boundaries, its processes and its control measures.

### 2.1.1. System Boundaries

The study area will often coincide with an administrative unit (region, district, province, etc.), because the administrative system usually requires an analysis of the functioning of the water resources within its administrative boundaries. The system boundaries of the NRS, however, depend on its physical characteristics. The NRS must include the administrative area, but may extend over a larger area, depending upon the physical boundaries.

In many studies it may be useful to subdivide the NRS into smaller units with suitable boundaries. Examples are subdivisions into a groundwater and a surface water system, subdivision of a surface water system into catchments and sub-catchments, and subdivision of a groundwater system into different aquifers or aquifer components. The definition of (sub)systems and their boundaries should be done in such a way that the transport of water across the boundaries can be determined properly.

### 2.1.2. Physical, Chemical and Biological Characteristics

The *physical processes* in a NRS are transport and storage within (sub)systems and transport between them. For the surface water system, a distinction is made between the infrastructure of rivers, canals, reservoirs and regulating structures (the open channel network) and the catchments draining to the open channel network.

The *chemical characteristics* define the composition of groundwater and surface water and the processes that may influence this composition, such as transport, degradation and adsorption. The level of detail to which chemical characteristics are considered has to be in line with the requirements and threats of water-using and water-related activities, the problems and possible measures to be considered, and the institutional setting.

A useful concept for describing the *biological characteristics* is the ecosystem. An ecosystem is a combination of abiotic and biotic elements that influence each other. It includes all the organisms in a given area and their interaction with the physical environment, so that a flow of energy leads to clearly defined trophic structure, biotic diversity and material cycles within the system. The prime driving force for the flow of energy and matter within the system is ultimately the energy from the sun. Through an often complex series of steps, the solar energy is used for production and reproduction, consumption and decomposition, succession and recolonization. Ecosystems are not static or 'mechanistic'. Rather, they are dynamic networks of interrelated parts, which – to a certain extent – are capable of auto-repair when some parts are lost. This dynamic character of nature has proven to be essential in evolutionary terms, as it provides enormous resilience in coping with physical disturbances such as volcanic eruptions and global climate change. With respect to the sustainable development of water resources, the dynamic character of ecosystems guarantees that human activities can achieve equilibrium with the environment, albeit at the expense of the quality of the ecosystem in terms of species diversity and the like.

In ecosystems, a distinction can be made between the aquatic, terrestrial and land–water sub-systems:

- The aquatic sub-system may be defined as all permanent water bodies: lakes, rivers, streams, reservoirs, ponds and so on. A characteristic ecological parameter is, for example, the flow velocity.
- The terrestrial sub-system may be defined as the area that is predominantly dry and receives water only from rainfall and groundwater. Soil characteristics, altitude, slope and climate are major parameters for its living conditions.
- The land–water sub-system (wetland) is the interface between the aquatic and terrestrial subsystems. It is characterized by a highly dynamic environment, caused by the frequent (often seasonal, but also day-to-day and annual) changes in the water level. Examples are the floodplains along rivers, reed belts fringing lakes and streams, intertidal areas (mangroves) and salt marshes.

In addition to this broad classification into three sub-systems, it is often appropriate to distinguish different landscape or vegetation units, such as habitats. As a first approximation, different habitat types can be identified on the basis of homogenous vegetation structure. Once the study area has been divided into a series of sub-systems, habitats or other ecosystem classes, it is important to assess the degree of spatial relationships between these subsystems. Ecosystems are rarely completely isolated from one another, but are linked by physical transport processes (e.g. water and soil transport), nutrient pathways and migration patterns, as discussed in Appendix A.

### 2.1.3. Control Variables: Possible Measures

Some of the system parameters described above can be used to control the system. By adding or changing system parameters, water resources managers can change the state of the NRS in ways they desire. An example is the rule curve describing the operation of a reservoir (when to release how much water for what purpose). Another example is the dimension of feeder canals. Increasing the dimension of these canals permits greater allocations of water to farmers. Control or decision variables can include those relating to the condition, design and operation of the WRS. An example of non-physical control that changes the state of the biotic system is the release of predator fish in reservoirs to reach a desired balance of species in the ecosystem.

## 2.2. System Characteristics of the Socio-Economic System

Like the NRS, the socio-economic system (SES) has its boundaries, processes, and control measures.

### 2.2.1. System Boundaries

The economic system generally does not have a physical boundary like that of the natural system. Activities in a river basin, for example, are connected to the surrounding economies through the exchange of goods and services. Also, from a social point of view, the boundaries of the natural and economic systems rarely coincide.

The factors that determine the socio-economic activities in a study area now and in the future should be analysed in the context of the problems being considered. They could relate to larger systems, such as the national or even the international economy, e.g. when prices on the international market are important for local agricultural activities. The boundary conditions for the socio-economic system are those factors that are beyond the control of the decision-makers. Examples of such boundary conditions are the state of the world economy, the value of the US dollar or the price of oil.

### 2.2.2. System Elements and System Parameters

The socio-economic part of the WRS can be defined by identifying the main water-using and water-related activities, the expected developments in the study area, and the parameters that determine these developments. Examples of activities or economic sectors that may be relevant and of the type of information that has to be obtained to be able to describe the socio-economic system are:

- Agriculture and fisheries: present practice, location and area of irrigated agriculture, desired and potential developments, water use efficiency and so on.
- Power production: existing and planned reservoirs and power stations, operation and capacity, future demand for electric energy.
- Public water supply: location of centres of population and industrial activities, expected growth, alternative resources.
- Recreation: nature and location, expected and desired development, water quality conditions.
- Navigation: conditions of the water depth in relevant parts of the open channel system.

- Nature conservation: location of valuable and vulnerable areas and their dependence on water quality and quantity.

Some examples of important system parameters of the SES are: population dynamics (life expectancy, birth rates, death rates, etc.), labour force and wage rates, price levels in relation to national and international markets, subsidies, efficiency of production and water use, and income distribution.

When identifying and analysing activities in the study area, it is important to consider possible discrepancies between the opinions of individual actors and their representatives. For example, individual farmers may have different interests than suggested by the official agricultural organizations.

#### 2.2.3. Control Variables: Possible Measures

The functioning of the SES can be influenced by legislative and regulatory measures, and the price of water may be a particularly important factor in deciding how much is demanded. This price can be influenced by the water resources managers and used as a control variable. When the cost of water use represents only a small portion of the total cost of an activity, however, an increase in its price may have little if any impact on water use. In some cases water use is a necessity of life no matter how high the costs, and in such cases, the price of water (or taxation for waste water discharges) is not a proper control variable.

### 2.3. System Characteristics of the Administrative and Institutional System

The administrative and institutional system (AIS), like the NRS and SES, has its elements that define its boundaries (its authority or limits) and its processes including its ways of reorganizing for improved performance.

#### 2.3.1. System Elements

To characterize the administrative and institutional setting, the responsible institutes at the national, regional and local levels have to be identified. In most countries, the following elements in the institutional framework can be distinguished:

- the central government, divided into sectors such as public works, irrigation, agriculture, forestry, environment, housing, industry, mining and transport
- a coordinating body, for example a national water board, to coordinate actions by various sectors of the national government
- regional bodies, based upon the normal subdivisions of government, for example provinces, districts, cities and villages.
- regional bodies, based on a division according to the physical characteristics of the area, such as river basin authorities
- water-user organizations, representing the interests of directly involved stakeholders, for example in irrigation systems.

The following information needs to be made available:

- which ministries and coordinating bodies have authority and responsibilities related to water resources management
- which agencies are involved in the preparation of water resources development plans
- which national and regional water resources development plans are available
- which authorities are responsible for implementing these plans (regulation, construction and operation of infrastructural works)
- what is the existing legislation (laws and regulations) concerning water rights, allocation of water resources, water quality control and the financial aspects of water resources management.

Other sources of information about the administrative and legal setting include the policy documents and plans of other water-related sectors such as environment, agriculture, economy, transportation, physical planning and energy.

#### 2.3.2. Control Variables: Possible Measures

From a systems point of view, the decision or control variables in the AIS are less clear than in the case of the NRS and SES. But in the case of the AIS too, measures can be taken to improve the functioning of the system, for example by establishing coordinating bodies when these are not present, shifting responsibilities towards lower levels of government, privatization and other measures.

## 3. Analytical Framework: Phases of Analysis

A water resources planning study generally comprises several phases. Although we do not suggest the use of a rigid framework, some distinct phases and activities can be recognized and used to structure the analysis as a logical sequence of steps. The description of these phases, the activities in them and the interactions between activities is referred to as the *analytical* (*or conceptual*) framework. A coherent set of models is used for the quantitative analysis of the water resources system, measures and strategies. This set of models and related databases will be referred to as the *computational framework*.

The purpose of a water resources planning analysis is to inform and support decisions. A decision process is not a simple linear sequence of steps, but involves feedbacks to earlier steps. Part of the process is thus iterative. A distinction is made between comprehension loops and feedback loops. A comprehension loop improves the understanding of a complex problem by cycling within or between steps. Feedback loops imply a return to previous phases of the process. They are needed when:

- solutions fail to meet criteria
- new insights change the perception of the problem and its solutions
- essential system components and links have been overlooked
- situations change (political, international, developments in society).

The general analytical framework for water resources management studies is depicted in Figure E.3. The three main phases of the framework are: *inception, development and selection*. Communication and interaction with the decision-makers are essential throughout the process. Regular reporting (inception and interim reports etc.) helps in effective communication, but a continuous dialogue is necessary at all stages of the analysis.

The first phase of the process is the *inception phase*. Here, the subject of the analysis (what is analysed under what conditions), the objectives (the desired results of the analysis) and constraints (its limitations) are specified. On the basis of this initial analysis, during which intensive communication with the decision-makers is essential, the approach for the

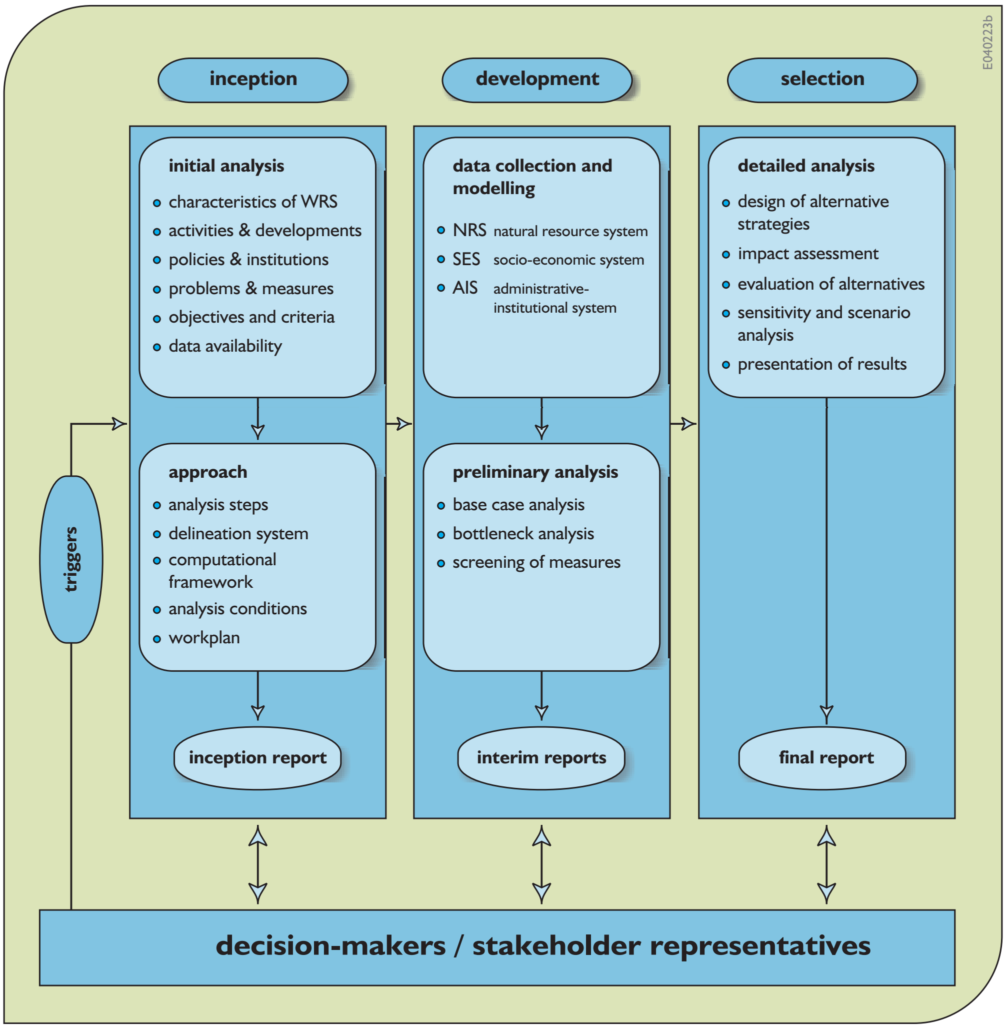



**Figure E.3.** Analytical framework for water resources planning studies.

remainder of the analysis is specified. The results of the inception phase are presented in the inception report, which includes the work plan for the other phases of the analysis.

In the *development phase*, the tools for the analysis and (elements of) solutions to the water resources problems are developed. Major activities are data collection and modelling. A preliminary analysis will ensure that the developed tools are suited to the generation and assessment of measures to solve water resources problems. Individual measures will be developed and screened in this phase. A gradual improvement of the understanding of various characteristics of the WRS is obtained, going from limited data sets and simple tools to more detailed data and the full set of models. Interaction with the decision-makers will be greatly enhanced if they are involved in the preliminary analysis as part of the analysis team. The formal interaction is structured through the presentation of results in interim reports.

The purpose of the *selection phase* is to generate a limited number of promising strategies that, after detailed assessment of their effects in terms of the evaluation criteria, will be presented to the decision-makers who – with inputs from the public – will decide on their preferred line of action. Important activities in this phase are strategy design, impact assessment, evaluation of strategies, sensitivity and scenario analysis, and presentation. The results of this phase are included in the final report, together with a summary of the results of the development phase. Continuous interaction with the decision-makers is essential in the phase. They or their representatives should be part of the team that carries out the analysis. During this phase, public consultations will also be needed to inform and involve stakeholders and get their reactions and inputs.

Reference is made above to the 'decision-makers'. Although it is clear that the analysis will eventually support decisions, it is not always clear who (or which unit) will make the final decisions. If an outside consultant performs the analysis, careful selection of the coordinating agency is essential for the successful implementation of the project. Interaction with the decision-making arena usually takes place through steering committees (which can act as an interdepartmental forum), technical advisory committees and public participation processes.

The inception phase is a very important component of this analytical framework. It determines to a large extent what will be done in the analysis. This inception phase should therefore get ample attention. Many commercial contract terms of reference (ToRs) for planning studies include an obligation for the 'consultant' to submit an inception report a few weeks after the starting date of the assignment. An inception report is simply an update of the proposal, covering the time between the submission of the proposal and the starting date, and addresses only the changes that need to be taken into account. Depending on the complexity of the system and required interaction with decision-makers and stakeholders, an estimate of the percentage of the total project resources (money and time) that should be devoted to the inception phase can amount to 30 or even 40%.

## 4. Inception Phase

Water resources planning studies are often *triggered* (left oval in Figure E.3) by specific management problems such as the need to increase power production, the occurrence of droughts or floods, or the threat of water quality deterioration. The need for water resources planning in relation to other sector planning efforts may also be a trigger. Which parts of the WRS are studied and under what conditions follows primarily from the objectives of the study (and from the available budget, information and time). The initiators of the study generally have more or less concrete ideas about the objectives and subject of the analysis. They react to triggers as mentioned above and, being aware of problems or issues, have taken the initiative for a planning analysis.

The client's ideas about the problems and issues to be addressed will usually be described in a Project Formulation Document (PFD) or ToR (terms of reference). The very first activity of the project is to review and discuss these documents. If the subject (what) and objectives (what for) of the analysis are adequately described in the ToR, the first step of the study is to specify the approach (how). In many situations however, the first task of the analysts is to assist the decision-makers in further specifying the objectives and subject of the analysis. For this activity, intensive communication is required with authorities involved in water resources planning and the stakeholders. They can provide information on the requirements of various interest groups related to water and on expected problems. It is not uncommon to have the stated objectives of a study differ from the actual (often unstated) objectives of the client. Furthermore, objectives change over time. As emphasized in Chapter 2, constant and effective communication between analysts and their clients is absolutely essential.

The inception phase starts with an *initial analysis*, to clarify the subject and the problems and objectives involved. Building on the understanding that developed from this initial analysis, a more or less formal description can be given of the *approach* that the planning study will follow, including a description of the analysis steps that will be taken, the computational framework that will be used and the assumptions and conditions under which the planning study will be undertaken.

## 4.1. Initial Analysis

The purpose of the initial analysis is to better understand the system (NRS, SES and AIS), its problems, management objectives and possible measures that can be taken to improve it (see Figure E.4). The analyst has to structure the activities and indicate the consequences of considering or neglecting various aspects. Brainstorming sessions with various specialists and stakeholders should help to identify and balance the different aspects of the study. The aim of these activities is to ensure that the analyst has the same perception of what needs to be done as the decision-maker and stakeholders.

### 4.1.1. Inventory of Characteristics, Developments and Policies

The initial analysis starts with an inventory of the *characteristics* of the WRS. This is not an easy task because it requires the reduction of a complex reality into a comprehensible description of system components. Choices have to be made about what (the detail that) should be included and what can be ignored. Such choices require engineering and economic skills in combination with a good understanding of the problems and possible measures.

The next step will be an inventory of the *activities and ongoing developments* that will determine how the system will function in the future and what kind of additional activities can be expected. This can include autonomous developments (such as population and urbanization growth) as well as policy decisions that have been or may be taken that will influence the WRS and/or its performance. An inventory of *policies and institutions* is needed to identify who is involved in the management and development of the system (and hence who should be involved in the analyses) and what their plans are.

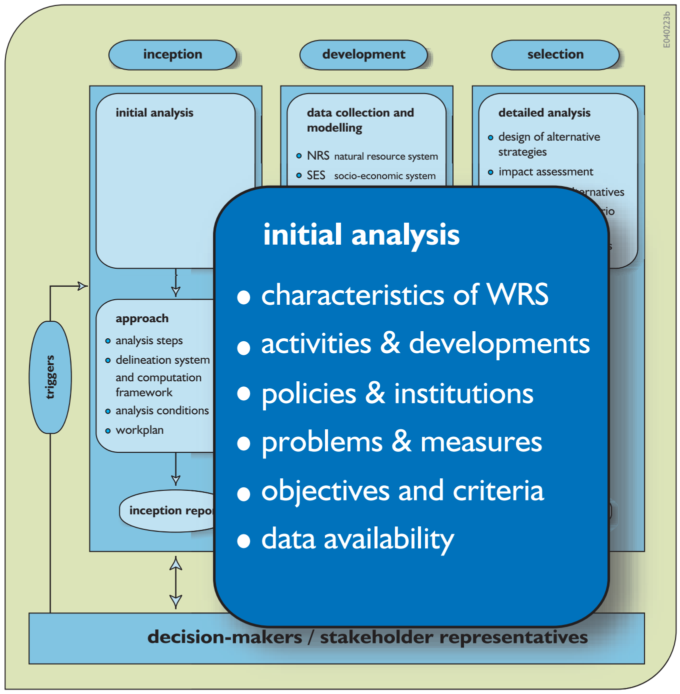



**Figure E.4.** Components of initial analysis.

This will contribute to the development of scenarios for the analyses. Scenarios are defined as developments exogenous to the water system under consideration (see Section 1.3).

### 4.1.2. Problem Analysis

Based upon the previous steps a problem analysis can be carried out. This should be done based on the facts and what the stakeholders and decision-makers perceive as a problem. A simple example of a problem analysis is presented below in Box E.2.

A problem analysis should be expressed as far as possible in terms of the socio-economic and environmental or ecosystem impacts that have meaning to the decision-makers. Not all stakeholders may be able to relate to predicted changes in flows, water levels or pollutant concentrations. Some may want to know how much money is involved (e.g. crop production, repair costs), the rate of shore line erosion, the relative change in fish population, or the number of people affected by flooding. Expressing outcomes in terms of socio-economic impacts makes it easier to relate the problems to the (socio-economic) development objectives that decision-makers have formulated for the particular region or system under consideration.

**Box E.2. Example of Problem Analysis**

A reservoir is in operation for power production and water supply to an agricultural area in a coastal zone. Farmers in the area are complaining that the water supply is inadequate. As the general objective for water resources planning is to develop the availability of water resources to obtain maximal economic benefits, an analysis is required.

A first analysis of the available information shows that the water demand for irrigation in August cannot be satisfied because the farmers have shifted to crops with a higher water consumption and the irrigated area has been extended. The constraint (bottleneck) is the capacity of the feeder canal to the area. If the capacity of the feeder canal is increased, the reservoir may become empty in dry years, unless the operation of the reservoir is modified. The use of groundwater could be considered as an alternative to surface water irrigation. Possible measures are:

- increasing the capacity of the feeder canal
- improving the operation of the reservoir
- increasing the withdrawal of groundwater from the coastal aquifer
- reducing the agricultural water demand in August.

A practical criterion to compare the effectiveness of the measures is their net benefit. Other factors, like the effect of groundwater withdrawal on saltwater intrusion or the loss of operational flexibility of the reservoir, may also have to be considered in the assessment of alternative solutions.

A good problem analysis will also indicate the measures that can be taken to eliminate, reduce or alleviate the identified problems. The problem analysis in Box E.2 includes such measures. The identification of measures does not only help to clarify the problems and possible solutions; such early identification is also needed for the design of the computational framework and the data collection activities (see Section 5.1). These activities should be designed in such a way that the measures can be evaluated in the analysis phases of the study.

### 4.1.3. Objectives and Criteria

An essential activity in the inception phase is the translation of general objectives, as described in policy documents, into operational objectives that can be

**Box E.3. Examples of Objectives and Criteria**

- General objectives:
  - to develop the potential of the water resources in such a way that the expected value of the (net) benefits to the national and regional economy will be maximized, while minimizing negative impacts on public health, welfare and the environment
  - to provide good drinking water to the population.
- Operational objectives:
  - to meet the given demands for public water supply (for population and industry) in a certain year
  - to realize a certain level of agricultural production in a given time
  - to meet certain predetermined water quality standards.
- Criteria:
  - cost-benefit ratio
  - amount (or value) of agricultural production
  - amount of energy production
  - area of aquifer lost to saltwater intrusion
  - changes in groundwater levels
  - environmental standards (BOD, DO, etc.).

assessed in a quantitative way. Some examples of objectives and criteria are listed in the Box E.3. The objectives and criteria used in a water resources management study in West Java, Indonesia are presented as an illustration at the end of this appendix.

To incorporate sustainability as an objective in the study, attention should be given to all aspects of sustainable development. The decision-makers and analysts must be aware of how sustainable development aspects can be incorporated in the water resources planning study. Criteria should be laid down to assess the sustainability of developments and measures that may contribute to sustainable development should be identified.

### 4.1.4. Data Availability

The last activity in the initial analysis is to investigate the availability of data and other information required for the study, in particular for the description of the NRS (quantity and quality) and the SES. The availability of data determines the level of detail and accuracy that can be achieved in the analysis. If few data are available, a more qualitative analysis may have to be performed.

The required level of detail will primarily be determined by what is needed to analyse the problems, objectives and measures involved. Data availability might be a constraint, but if abundant data are available they need not be all used (by applying very detailed computer models, for example). The relevant problems and measures might very well be addressed by a less detailed approach or even a qualitative analysis.

On completion of the initial analysis, the analyst should have a clear idea about what will be studied, for what purpose and under what conditions. The information related to those questions will be documented in the second part of the inception phase, the phase that defines the scope of the study.

## 4.2. Specification of the Approach

Once it is clear 'what' will be analysed and 'why', it is the task of the analysts to specify 'how' this will be done. Figure E.5 highlights the elements involved. The specification starts with the steps that will be followed in the analysis, enabling the stakeholders to follow the process. Next the system will be described, laying out what will be analysed and what will not. To assess the effects of measures or combinations of measures, a computational framework is required. The analysis conditions specify the assumptions and conditions under which the analysis will be performed. All required activities will be combined in a work plan. Some important elements in the specification of the approach are discussed below.

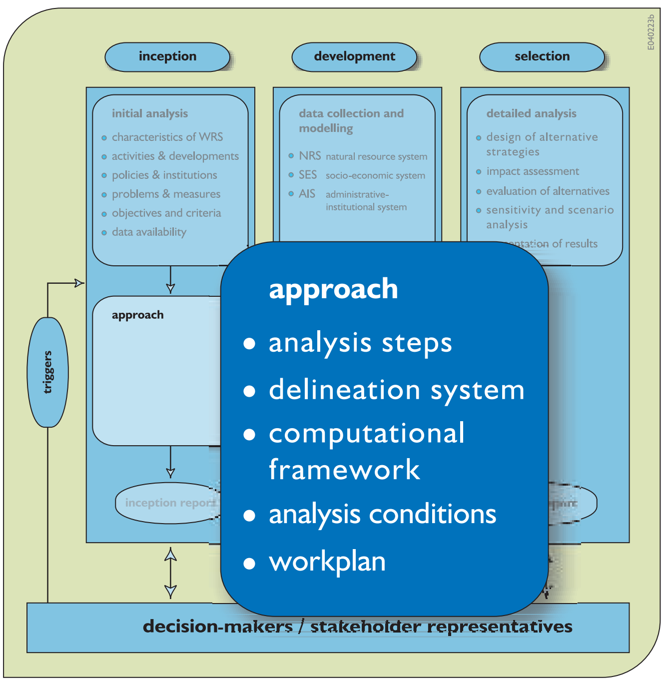



**Figure E.5. Components of the inception approach.**

### 4.2.1. Analysis Steps

The specification of the analysis steps has both an internal (to the project) and an external function. The *internal function* is that it will tell the various (often interdisciplinary) team members what the position of their contribution is in the overall analysis approach. This is necessary to schedule activities of all the members and to ensure (or at least enhance the probability) that they will provide the right kind of information to each other at the right times. The external function relates to the interaction with the decision-makers and stakeholders; it specifies when and how they will be involved in the analysis.

The specification of the analysis steps can be based on Figure E.3. This generic figure should be made more specific by addressing the particular aspects of the system and problems under consideration.

### 4.2.2. Delineation of System

The next step is to determine the components that will be taken into account in the analysis. The boundaries of the (sub)systems need to be determined on the basis of their natural boundaries, physical in case of the NRS, administrative in case of AIS, etc.

#### Required Level of Detail

An often difficult aspect to address in the inception phase is the level of detail that should be chosen for the components and subsystems of the system under study. To comply with the overall purpose of the study – the preparation of a plan that addresses the problem and provides 'the best use' of goods and services provided by the WRS – one does not have to know the use of every square metre of land in the study area or the water use of each leather tannery in the area. One has to use values averaged over areas and groups of users with just enough detail to provide the information required to support decision-making. The information is used to discriminate between alternative courses of actions and has to be relevant for the evaluation process. It should thus be expressed in terms of the selected evaluation criteria.

One of the main tasks of a project leader is to manage the experts from various disciplines in their 'enthusiasm' to find the answers to all possible questions. Not staying focused on the appropriate level of detail is one of the most common causes for project failure. If the needed level of detail is underestimated at the start of the project, the study has to go into more detail to be able to solve the problem. Sometimes the right level of detail is chosen, but team members spend too much time addressing more detailed questions and fail to come up with answers within the available time. Changing the level of detail is one of the main reasons for feedback loops in the analysis process.

The level of detail has to be specified for the following aspects:

- *Spatial disaggregation of the study area: type and size of units*. The spatial disaggregation usually depends on the types of water demand, the system of rivers and canals, and the characteristics of groundwater aquifers.
- *Sectoral disaggregation of water users*. The ISIC classification (International Standard Industrial Classification) distinguishes economic activities in various sectors. At the two-digit level for example, agriculture is treated as one sector. At more detailed levels the agricultural sector is broken down in smaller subsectors such as aquaculture, dairy farming and cotton growing.
- *Time steps for the analysis of processes*. It is important to select the most appropriate time scale. One generally has to consider various time steps. In a system with reservoirs with over-year storage, one should run a sequence of years to study the performance of the reservoir. For the analysis of the agricultural sector one often has to account for seasonal variability.
- *Quality of the environment*. Not all species can be used as indicator species. A choice has to be made to show the impacts of measures on the quality of the environment and the ecological integrity.

### 4.2.3. Computational Framework

The results of the preliminary analysis will define the computational framework that is needed for the analysis. This framework comprises mathematical models, databases, GIS and the like. These must describe the system (NRS, SES) and evaluate possible measures and strategies under different scenarios. (The different kinds of models and approaches that can be used are described in the main text of this book.) In general, a combination of simulation and optimization models will be useful.

For the development of the computational framework, the study area may be subdivided into smaller units considered to be homogeneous with respect to their characteristic parameters. Each unit is represented in the mathematical model(s) by a computation element. The number of elements required for the analysis depends on the issues being addressed, the complexity of the study area, the kind of measures to be studied and the availability of data. It generally is wise to start with a preliminary schematization with the minimum number of elements. If more spatial detail is required in a later stage, elements can be subdivided.

#### Use/Adaptation/Development of Computer Programs

Computer programs may be used to analyse various aspects of the WRS. A decision should be made about whether to use existing programs or to adapt or develop new ones. As development of computer programs is a very time consuming activity, existing ones should be used whenever they meet the requirements of the study.

#### Hydrometeorological Boundary Conditions

Boundary conditions for models of the NRS comprise precipitation, evapotranspiration and flow across the model boundaries. Parts of the NRS that are not influenced by water management (natural catchments) are usually not included in the computational framework. The discharges of these (sub)catchments are derived from historical streamflow data. If the availability of streamflow data is insufficient, rainfall–runoff models may be used to generate streamflow records from precipitation data.

Three approaches may be used to account for the stochastic character of the hydrometeorological conditions for which water resources strategies have to be evaluated:

a. Simulation of a strategy for a complete set of hydrometeorological data, e.g. forty consecutive years, including dry and wet periods.
b. Simulation for a selected sample of the available data, each one representative for a certain situation, such as an average year, a 10% dry year or a 2% dry year.
c Simulation of just one representative year, such as an average year.

Which approach is preferred will depend on the specific problem situation, variability of the hydrometeorological conditions, availability of data and the specific analysis that is required. It is clear that in cases where substantial over-year storage in reservoirs (surface water and groundwater) is involved, approaches b) and c) are not viable. Approach c) is only applicable in situations with very little variability over the years. Approach b) is often a good compromise. Figure E.6 illustrates the approach. The simulations are performed for three representative years: an average year ($D_{50}$), a moderate dry year ($D_{10}$) and an extreme dry year ($D_5$). The expected damage in this example is approximated with the following formula:

$$\text{Expected damage} = 0.125 * D_5 + (0.350 - 0.125) * D_{20} + (0.650 - 0.350) * D_{50}$$

In the development phase, when many alternative measures (solutions) have to be screened, simulations will be made for selected conditions only. For screening (see Section 5.2.3), the system is generally studied under stress conditions, for example a 10% dry year. The minimum time interval taken to represent a certain situation depends on the memory of the system. In a system comprising reservoirs with considerable over-year storage, representative periods should be more than a single year. In the selection phase, simulations will be made for promising strategies only, so that the complete set of historical data can be applied to them.

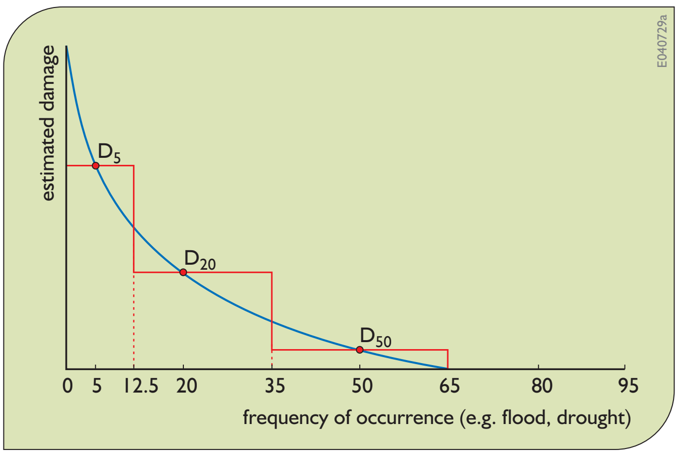



**Figure E.6.** Expected damage calculation based on three representative years.

### Requirements for Data Collection

The analytical approach described above is usually formulated in an iterative way, taking into account data requirements, data availability and available resources (budget, manpower, models and analytical techniques). When the approach has to be modified in view of constraints (usually budgetary), the data requirements are reassessed. Once the analytical approach is specified, the exact requirements for data collection can be formulated: which data need to be collected, when, and by whom. Opportunities to collect additional data from field observations will generally be limited by the available budget and time.

### 4.2.4. Analysis Conditions

The assumptions and conditions under which the analysis is undertaken have to be specified in close cooperation with the decision-makers. These assumptions and conditions include:

- (Sub)systems boundaries. Definition of the WRS and a possible subdivision in sub-systems are based on physical (NRS) and/or administrative boundaries (AIS, see also Section 4.2.2).
- Base year. The analysis will be made for the present condition and one or more points in time. The present situation may be defined by the situation in the most recent year for which a reasonably complete set of data is available (the base year).
- Time horizons. These will be selected on the basis of existing planning horizons – usually ten and twenty-five years – or be related to the characteristic time scales of subsystems.

- Scenario assumptions concerning factors external to the WRS, such as the growth of population, food consumption and of water-related economic sectors and prices (energy or crop prices, for example). In many cases an analysis of socio-economic data will be required to obtain reasonable ranges for scenario variables (see Section 5.1.2).
- System assumptions. These concern factors internal to the WRS, such as the response of crop production to improved cultivation practices, or the effectiveness of price incentives on per capita water consumption. These system assumptions can be subject of additional (sensitivity) analysis in the selection phase (see Section 6.3).
- Time and budget constraints. The study has to be executed within constraints of available time and budget.
- Data availability. The study must be based on the data that can be made available within the time and budget available.

The choice of the time horizon is often given insufficient attention. Official planning horizons (e.g. ten and twenty-five years) are often used as time horizons for all elements of the analysis. As well as the planning horizon of the traditional government planning framework, however, one should also consider the time scales of the system and the processes within it. System components will have characteristic time scales. For example:

- User functions (economic activities) have life cycles that are determined by the amortization period of the investments. The time horizon of the planning process is based on these conditions.
- Social institutions have time horizons that depend on the pace of legal/institutional and political decision-making.
- Physical–chemical systems have time scales that depend on the response or restoration times of the systems. Restoration of polluted rivers, for example, may be achieved within a few months, while the restoration of a polluted groundwater aquifer may take decades.
- Ecosystems may have a time scale of a few weeks algae blooms) or tens of years (degradation of mangrove forests), depending on the type of process or intervention.

To study the sustainability and ecological integrity of the resource system, time horizons should be tuned to the response times of the system rather than to a planning horizon only. Although more attention is now paid to sustainability, no operational procedure has been developed to give long-term effects their proper place in the evaluation process. Decision-makers tend to take short-range decisions, involving possible risks in the long term, because their political jobs are often limited to (or renewable in) short terms and hence they prefer short-term political gains.

### 4.2.5. Work Plan

The results of the inception phase are documented in an inception report (Figure E.3), which will be used as a reference during the execution of the study. An essential part of the report is the work plan, in which time, budget and human resource allocations to various activities are specified. This work plan typically includes bar charts for activities and staffing, time schedules for deliverables, milestones, reporting procedures and similar features. It should include a communication plan that describes the interaction between the decision-makers and stakeholders and the analysis team.

## 4.3. Inception Report

The inception report has to contain all the results of the inception phase. It should make clear what will be studied, why and how. In many cases it will also specify what will not be studied and why. The content of the inception report follows the subjects mentioned above under the initial analysis and approach:

- Description of the water resources system
  - NRS: components
  - SES: activities and developments
  - AIS: institutions involved, legal aspects including a delineation of the components that will be studied.
- Problem analysis and problem statement.
- Objectives and criteria.
- Analytical approach (steps, models and databases).
- Analysis conditions (time horizon, scenarios, assumptions and so on).
- Work plan.

### 4.4. Communication with Decision-Makers and Stakeholders

The inception report is a specific and concrete result of the inception phase. It is an important product because it contains all that has been learned in this first phase and that has been agreed upon between the analyst and the 'client' (the decision-makers and the stakeholders). A possibly even more important result of the inception phase, however, is the interaction that should take place during this phase between the analyst and the client. The client's views about problems, objectives and other aspects should develop alongside the view of the analyst. The analyst must understand the client's concerns, problems and objectives and not just his or her own. Clients should feel they 'own' the results of the inception phase and view the inception report as their own product, not merely a report of the analysts or consultants.

To achieve such ownership, frequent interaction must take place among the analysts, the decision-makers and stakeholders, to a much greater extent than is indicated in Figure E.3. This can be done in specific workshops, such as those devoted to the problem statement or to the specification of objectives and criteria, public consultation meetings.

## 5. Development Phase

The development phase includes simultaneous model development and data collection followed by a preliminary analysis to identify possible solutions for the problems being addressed.

### 5.1. Model Development and Data Collection

In the development phase, a coherent set of computational tools is developed for the analysis of the WRS and SES, as specified in the inception phase. Possible solutions for the water resources problems are also identified in this phase. The result of the data collection and modelling activity is a computational framework that represents the characteristics of the WRS at an adequate level of detail. The framework is designed to quantify the effects of individual measures or combinations of measures, expressed in values for the evaluation criteria that correspond to the objectives of water resources management. For the schematization of the study area, experience from previous applications may be helpful. However, each study is different and the schematization should always be tailored to its specific requirements.

If computer programs have to be developed from scratch, or if existing computer programs have to be adapted in a significant way, a considerable effort may be required and the activity may consume a large part of the available budget and time. Careful selection of the phenomena to be represented by the models, tuned to the needs of the analysis, can result in considerable savings. In the course of the modelling activity, more information on the study area and the type of measures to be considered may become available. Sometimes this may lead to a simplification of the modelling activity. The models should therefore be flexible and adaptable to new information. Activities related to data collection andmodelling will comprise all three components of the WRS (see Figure E.7). The emphasis is generally on the NRS, and to analyse this, a distinction can be made between physical, chemical and biotic components.

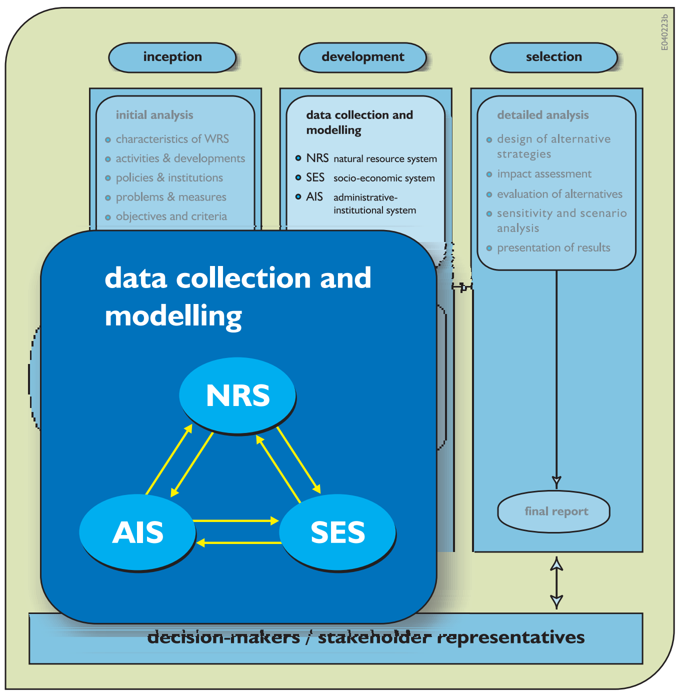



Figure E.7. Components of data collection and analysis.

### 5.1.1. Analysis of the Natural Resources System (NRS)

#### Physical Components

The NRS comprises the natural and human-made infrastructure, including the hydrometeorological boundary conditions. Models can be used to simulate the processes of water distribution through the infrastructure, taking into account the storage of water and withdrawal from the network to satisfy the demands of the water-using activities. (Chapter 11 provides an overview of possible models involved.)

The main water allocation related functions of the surface water system are storage of water in reservoirs and diversion of water from the rivers and main canals to branch canals. Typical time steps in planning studies are periods of a week, ten days, fifteen days or a month. In most river basin simulation models (e.g. RIBASIM, WEAP, MIKEBASIN), the surface water system is represented by a network of nodes and branches. At the nodes, natural runoff can be supplied to the system or water can be abstracted to satisfy specified water demands. Storage is represented by reservoir nodes, where reservoirs are operated according to user-specified rule curves. Facilities are available to account for power generation, diversion of water to branch canals, return flow from irrigated agriculture, drainage from districts, minimum flow requirements and so on. The allocation of water to the demand nodes is in upstream–downstream sequence, accounting for user-specified priorities. Most river basin models contain algorithms to include simple flow routing in the analysis. If necessary, more detailed hydrodynamic models such as MIKE-11 or SOBEK can be used.

Hydrometeorological data are analysed to provide time series of inflows to the surface water system. Typical elements of the analysis are: historical streamflow data, rainfall data and rainfall–runoff relations. The reliability and availability of historical streamflow data are often insufficient for the purpose of the study. Streamflow data may then have to be supplemented with synthetic data derived from observed precipitation and rainfall–runoff relations. A typical period for which data would be required is twenty-five years. Chapters 7, 11, and 13 introduce tools for an analysis of hydrometeorological data and rainfall–runoff simulations.

A main parameter characterizing the groundwater system is the groundwater level or head. In coastal zones, where seawater intrusion may occur, the salinity of the groundwater has to be considered as well. The time scale for effects of withdrawals on the groundwater head is in the order of ten to fifty years; saltwater intrusion is even slower and effects may have a time scale of a century or more. For such computations, steady flow conditions are often assumed and a comparison is made of the effects of different abstraction alternatives. Important aspects in the analysis of the groundwater system are the identification of the system of aquifers and aquitards and the determination of the recharge to the aquifers. Groundwater models can be used to test the consistency of collected data. Use of models to quantify the effects of future groundwater abstraction is only suggested in cases where there is sufficient information to calibrate the models for reproduction of the existing situation. Many general groundwater programs (e.g. MODFLOW) are available to compute the groundwater heads in a system of aquifers.

#### Chemical Components

The analysis of chemical components in the water system is used to study the influence they have on the user functions or the biological system. The components and processes that are to be considered in the analysis will have been selected in the inception phase. Their concentrations are computed in discrete segments of the surface or groundwater system, and this will require data on the influx of substances across segment boundaries and on the chemical processes in the segments. (Chapter 12 gives more detailed descriptions of available modelling approaches.)

The results of the water quantity modelling may be the inputs for water quality models, and are used to determine the water balance for each segment. To assess the mass balance of the chemical components, data on concentrations in the inflows to the segments are required. In addition to the inflow terms in the water balance, some extra terms have to be specified. For example, the discharge of sewage water may be negligible for the overall water balance but have a significant impact on the water quality of a river segment.

#### Biotic Components

The analysis of the biological system aims to determine the response of the ecosystems to water resources management. Because there is too little exact information on the biotic components in most cases, and because ecosystems largely depend on habitat parameters, the analysis may have to be limited to the essential relationships between ecosystems and physical and chemical parameters. The selection of key factors for ecosystem performance depends on the type of ecosystem, the kind of analysis for which they will be used and the type(s) of activities or human interventions involved.

For example, consider the impacts of dam construction on downstream ecosystems. Key parameters include:

- hydrodynamic regime (discharge characteristics and flooding pattern, such as duration, depth and frequency)
- water quality (e.g. sediment load, temperature, oxygen, salinity).

It is likely that these parameters will change due to the dam's construction, which may thus affect the downstream ecosystems. A combination of information on the present ecosystems and physical conditions and the direction and order of magnitude of change in the parameters mentioned above often enables one to make an expert judgment on the significance and direction of change of the ecosystems.

However, if one wants to make a more accurate and detailed prediction of ecosystem changes, for example by using mathematical modelling, much more information on the conditions determining fluvial ecosystems is needed. For example, a vegetation development model of the downstream areas liable to flooding needs – in addition to the key parameters mentioned above – information about:

- local topography (relief, dikes and culverts, channels, oxbows etc.)
- local geomorphology or soil types
- vegetation patterns.

Using known relationships between vegetation and physical factors such as the inundation frequency and soil/geomorphology, a preliminary model can be made that predicts changes in the vegetation pattern caused by changes in river flooding behaviour. Such an approach can be greatly enhanced by the use of geographically oriented data, in the form of digitized maps of the topography of the area, the soils, inundation and vegetation.

For aquatic habitat types, the approach may be quite different. Aquatic flora and fauna in rivers and lakes are affected by such factors as flow velocity, shear stress, water depth, bottom type and water quality. The use of one-dimensional hydraulic and water quality models is therefore often indispensable, but needs a concerted effort between ecologic and hydraulic modelling experts. It is especially important to identify data requirements as early as possible. In most cases, special physical or chemical modelling calculations are needed to supply the required data for habitat prediction. For example, hydraulic modelling often concentrates on high and low water levels, whereas floodplain ecosystems also depend on intermediate levels and duration curves. Time steps may differ and additional detail may be required in the model.

In most cases, predictions of ecosystem changes have a relatively large margin of uncertainty. Causes of this uncertainty include the often limited availability of data and the gaps in our fundamental knowledge of how ecosystems really work and of their complexity and dynamic behaviour. Creating detailed descriptions of species composition remains very difficult, especially when new species may appear. These factors also make it very difficult to assess the significance of the changes. The evaluation of ecosystems generally requires information on species diversity and the occurrence of rare or endangered species, and this is often lacking.

The significance of the effects of water management on ecosystems is increasingly acknowledged. Specific tools for predicting these effects are being developed, mostly on a regional basis. For example, models can be used for the prediction of vegetation development in relation to inundation characteristics, soil type and nature management practice in floodplains, or in relation to groundwater management. For biological water quality, models are widely available, including models for algal composition and algal bloom (Chapter 12).

## Scenario Development

Depending on the scenarios defined in the inception phase, it might be necessary to develop scenarios for the natural resource system. In nearly all cases, these willcertainly be needed to take into account possible climate change that may affect the physical components (rainfall, flow, etc.), and very probably also the chemical and biotic components.

Scenarios for the natural system are also needed when the system boundaries, as defined for the analysis, do not cover the full river basin and, for example, developments upstream might influence the inflow into the system. This is typically the case in national studies where the system boundary has been set at the national border.

### 5.1.2. Analysis of the Socio-Economic System (SES)

Developments in the SES determine the way demands on the NRS will develop. Conversely, the development of economic activities within the study area may depend on the availability of water. For example, good supplies of relatively cheap surface water may stimulate the development of irrigated agriculture, or attract industrial activities that require large quantities of water for their production processes. Another example is the development of water-based recreation around a reservoir. These socio-economic system developments in turn increase the water demands.

It is the task of the economist or planner to estimate the future levels of the activities that require water resources for their development, as well as the resulting water demands (and discharges). The estimate is made for the time horizon of the study, within the geographical boundaries of the area. The main activities are:

- assessment of the present economic situation
- activity analysis
- scenario specification.

#### Assessment of the Present Economic Situation

The starting point for an analysis of the SES is an assessment of the present economic situation with respect to the water-related activities and the factors that determine these activities. Past trends should be analysed to provide information on factors that have been decisive in bringing about the present situation and that may give clues about future developments. The analysis should focus on present and future major water use categories, which are identified at an early stage in the analysis, rather than trying to estimate water demands and discharges for all possible water use categories.

One's attention, then, should be on the most important factors that determine relevant water-related activities rather than on an analysis of the total economy. The difficulty in forecasting economic developments is, however, that one does not know *a priori* which factors will be decisive for these developments. As population dynamics are explanatory factors for many economic developments, demographic analysis (population growth, where and at what rate) is generally one of the first activities in this phase. Other important parameters for the development of water-related activities are:

- sectoral distribution (primary, secondary and tertiary sectors)
- industrialization
- labour force and wage development
- the price of oil and other commodities
- the balance of trade (imports and exports)
- the stability of local currency.

In the analysis of the present economic situation, attention should be given to the possible effects of changes in the WRS on the regional or national economy. To determine how effects related to water users may influence major economic indicators such as gross national product, balance of trade or equity considerations, a macro-economic analysis of the relation between the water-related sectors and the rest of the economy and its surroundings (e.g. foreign trade) is often required.

An example of a macro-economic model that could be used for such an analysis is an input–output model. Such a model shows the relationships between various sectors of the economy. Using an input–output model, one can analyse how effects generated in one sector influence other sectors of the economy, not only in terms of value added, but also in terms of labour requirements. Such economic models have been extended to include environmental and resource-use aspects (national resource accounting models). Inclusion of these aspects, however, requires a rather large effort (in terms both of data and of manpower) and should therefore only be considered when major effects on the national economy are expected.

#### Activity Analysis

The activity analysis focuses on the relation between the economic activities and their water use. It focuses on the relations that determine the type and amount of water used by various activities. An activity analysis must answer the following questions with respect to each identified activity:

- What are the amounts of water (quantity and quality) demanded during which periods of the year and at which locations?

- What are the amounts of water discharged and the pollution loads during which periods of the year and at which locations?
- What are the benefits to the user if these amounts are made available?
- What is the damage to the user if these amounts cannot be made available?
- Which costs can be recovered by having the user pay for the water?
- Will cost recovery influence the water use pattern of the water user (both at the intake and the discharge side of his activity)?

Estimates of future water demands and wastewater discharges are generally based on the determination of unit water use and wastewater discharge per unit of activity. Looking at combined trends in unit water demands and economic developments can provide insight in possible future water demands. The final result of the activity analysis should be water demand functions, relating water demand to exogenous parameters.

As well as the level of activities and the resulting water demands, the geographical location of water-using activities (the pattern of activities) must also be known or estimated. If the pattern of activities is not expected to change, the analysis can be focused on the present situation in the study area. If new activities are expected to develop within the study area, it may be necessary to analyse the water-use characteristics of similar activities in other regions.

The resulting water demands should not be considered as 'given'. Water-use coefficients can be changed through measures such as water pricing that aim at reaching a socially preferred use pattern. Technological developments may result in less water use and pollution load per unit of product. If supplies and demands are matched before the effects of such incentives are analysed then one may end up with an overcapacity, because the 'given' demands may be lower if water users are confronted with the costs of the use of the water resources. This type of feedback needs to be considered in the study.

Estimating benefits and costs related to water use provides information to evaluate water resources management strategies. One of the criteria in this evaluation is the net economic benefit of the measures for various water users. One needs to know what benefits water users derive from the use of the water resources, and what costs they have to pay in order to obtain these benefits. The activity analysis should produce therefore benefit and cost functions related to various water uses. These functions can be used to evaluate the proposed strategies.

## Scenario Specification

Because of the multitude of factors that determine water demand, perfect forecasting of economic development and resulting water demands is a utopian dream. Future water demands are often dependent on future scenarios. A water demand scenario is a logical combination of basic parameters of the economy, their effects on water-related activities, and the resulting water demands. An understanding of the functioning of the socio-economic system developed through the assessment of past and present trends should be used to formulate a limited number of consistent scenarios.

For analytical purposes, it is useful to define an autonomous development scenario (i.e. the development of the socio-economic system) and its water demands, as it would develop independently from changes in the WRS. An autonomous development scenario does not consider possible constraints in the supply of water, and is thus very useful for analysing bottlenecks in the water resources system.

In a regional water resources management analysis, national economic growth and technological development are usually taken as exogenous parameters for the autonomous development of the region. Regional measures can however influence water use and may divert economic activities away from the area of study, or even influence the growth of the national economy, which was originally seen as an exogenous parameter.

Since many uncertain factors influence the development of water demands, the future cannot be projected by using a single autonomous development scenario. Rather than constructing many alternative scenarios through permutation of the uncertain parameter values, one may have to consider only two or three different development paths.

Scenario building is not a straightforward activity; it is partly an art that depends largely on the views of the analysts and local experts about which trends are likely to prevail in the future. The developed scenarios have to be internally consistent with respect to the magnitude and direction of the developments. Exogenous variables (population growth, economic growth) and water-related

**Box E.4. Example**

The development of water demand in agriculture is autonomous only to a limited extent, as it depends largely on the availability of land and water resources. The demand for agricultural products, however, will develop in an autonomous way. If the availability of water resources in a region is limited, the autonomous development of agriculture will be limited as well, and one would predict a small increase in agricultural water demand. If the demand for agricultural products increases considerably and self-sufficiency in food production is an objective, then the desired agricultural development to meet this objective may be considerable. The water demand corresponding to this desired agricultural development will show the need for further development of the water resources in the region.

variables (for example per capita water consumption) are combined into logical, coherent water-demand scenarios. An increase in per capita income could cause an increase in per capita water consumption. A scenario which presumes high economic growth should therefore also presume high per capita water-use coefficients.

Scenarios are often made top-down: that is, first one projects the basic parameters of an economy (population, gross national product, imports and exports, etc.) for the whole country or large parts of it. The economic developments projected at the national level are then broken down for regions, usually using past trends in regional economic growth related to the national average growth trend. These regional projections should be in line with population projections, because economic development also depends on population growth. One should use information available at the local level to check the assumptions made in disaggregating national projections.

### 5.1.3. Analysis of the Administrative and Institutional System (AIS)

An analysis of the AIS is required for three reasons:

- The information required for the study has to be obtained from the agencies involved.
- The legal framework may impose certain constraints on the development of water-using activities.
- The implementation of measures may be constrained by institutional bottlenecks.

In the inception phase, a preliminary analysis is made of the institutional setting of water resources management. The responsible authorities and institutes in the water resources sector and water-related sectors are identified at the national, regional and local level. In the analysis phase, the responsibilities and instruments of the authorities are reviewed. Attention must be given to the interaction between various authorities involved in water resources management and to the effectiveness of the AIS.

The responsibilities and authorities of various agencies are generally described in national legislation. An analysis of the legal framework should identify the various agencies responsible for (aspects of) water resources managment, their authority, their instruments and their coordination with other agencies. Arrangements made in the past concerning the use of water (water rights) should be carefully studied, since these may significantly constrain the options for water resources development.

Water resources management studies are often limited to the preparation of policies for a certain agency. In this situation, the analysis of the AIS will mainly serve to identify measures that can be implemented effectively by that agency. The responsible agency should be aware of the possible solutions and the role they may have in solving the managementproblem. Sometimes, the analysis of the AIS may result in recommendations for institutional and legal changes.

### 5.1.4. Integration into a Computational Framework

The various models and components developed for the NRS and SES describe parts of the total system. Some models may produce output that is needed as input for another model. Some may run interactively. This means that all models will have to be linked in a computational framework. Depending on the models and the problem situation, such linkage can be very strict and formalized in a kind of decision support system. In other cases, a clear description of information flow from one model to the other may be sufficient. In optimization models the linkage of the various components is taken care of by the structure of the model.

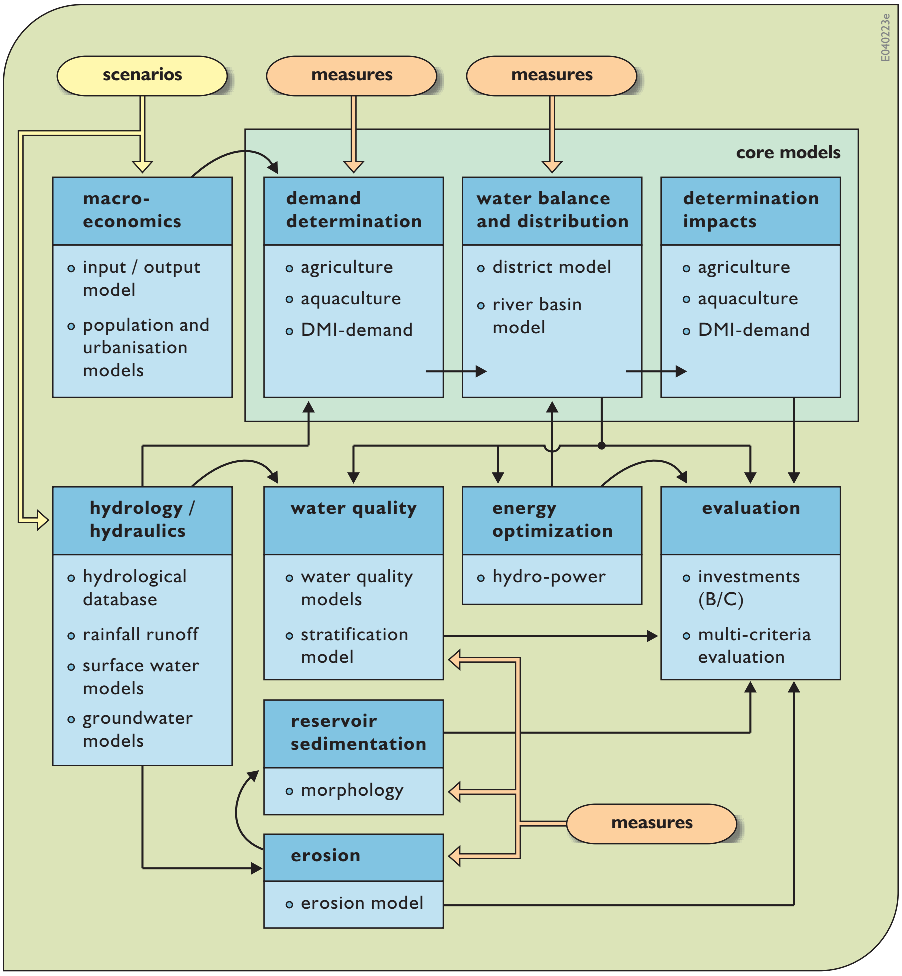



**Figure E.8.** Typical computational framework of simulation models.

Figure E.8 gives an example of a computational framework in which various simulation models are combined to analyse a river basin under drought conditions. The reservoirs in the system require special attention for the quality of the water they contain, sedimentation and the optimization of hydropower generation. The core of this computational framework is formed by the 'core models' block in the upper right corner of Figure E.8. In this block the demand for water is determined, followed by a balancing of supply and demand and the determination of the implications of the water allocation for the use functions. The models and modules in this core will typically be combined into one unit with automated linkages between the models and modules. The other models are linked through file transfer. This applies to the required input on macro-economic and hydrometeorological conditions (generated by scenarios) as well as the side analysis with respect to the modelling of the sedimentation and water quality in the reservoirs. The last parts of the computational framework are the modules that determine the financial and economic aspects (investments, operation and maintenance, benefit–cost, etc.) and support the overall multi-criteria analysis. At various places in this computational framework, measures can be included by changing input parameters. Scenarios can be analysed by changing the macro-economic and hydrometeorological conditions.

Figure E.8 is just an example. Other problem situations may require different computational frameworks. Computer modellers and analysts have a tendency to make these computational frameworks rather complicated and fancy. Sometimes this complexity is necessary, but in most cases a simple computational framework will be sufficient and preferable. The basic approach is to start as simple as possible and only add details when they prove necessary to carry out a proper analysis.

## 5.2. Preliminary Analysis

The 'preliminary analysis' includes the application of the knowledge obtained from data collection and modelling (Figure E.9) to identify possible solutions for the water resources management problem. At the start of the development phase, the analysis is focused on problems in the present situation (base case analysis) and the future situation (reference or bottleneck analysis). Knowing the problems that are to be expected under certain conditions, measures to solve them can be defined.

Water resources development should be guided by the principle of sustainable development. It should lead to a sustainable economy, in which natural renewable resources are used without being depleted (in either their carrying capacity or environmental utilization space). Sustainable economic development can also be supported by the development and adaptation of knowledge and improvement of organizational and technical efficiency.

Many alternative measures can be identified in general, each making a contribution to the (partial) solution of a particular problem. To avoid spending time on less effective measures, the effects of individual measures should be compared and promising measures selected for further analysis. Preliminary versions of the computational framework can be used to get a first idea of the effects of different measures. During the development phase, when it becomes clear which types of measure are promising and which elements in the computational framework are critical for the assessment of their effects, the computational framework can be modified.

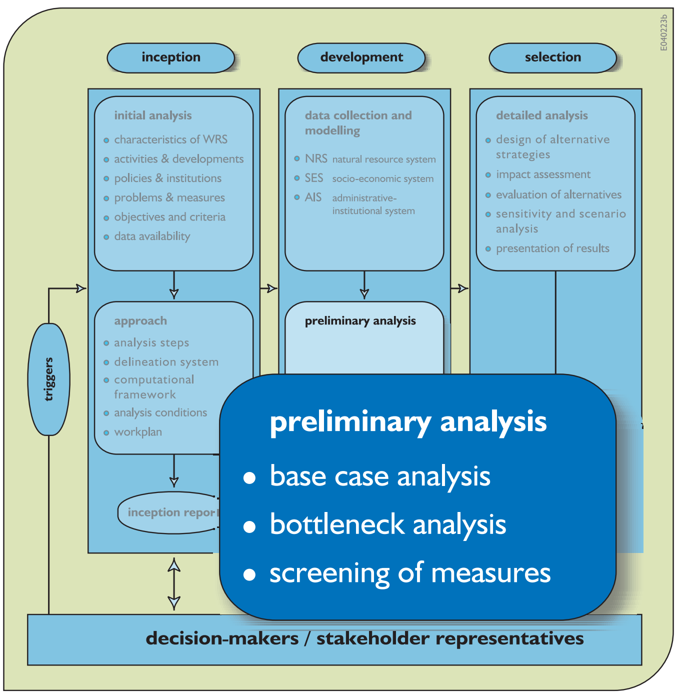



**Figure E.9.** Overview of activities in the preliminary analysis.

**Box E.5. Definitions**

- Base year: present situation.
- Time horizon: future situation.
- Base case: performance of WRS in present situation.
- Reference case: performance of WRS in future situation if no additional measures are taken.

### 5.2.1. Base Case Analysis

In the inception phase, a provisional problem statement is formulated on the basis of the results of the initial analysis. In the development phase, computational tools are used to further elaborate the problem statement. This is done by specifying two cases: the base case and the reference case.

In the base case analysis, the performance of the WRS is studied for the infrastructure and water demands in the base year, which is the most recent year for which a complete set of data can be collected. The base case describes the performance of the WRS in the present situation. A comparison of this base case with targets as specified in the management objectives produces a quantified problem statement.

From a modelling point of view the base case analysis also functions as a kind of calibration of the models. The results can be compared with measured information and, when necessary, parameters can be adjusted to better reflect the reality. Sufficient attention should be given to proper calibration and validation procedures.

### 5.2.2. Bottleneck (Reference Case) Analysis

An essential activity is an analysis of the WRS's performance if present policies are continued ('no change alternative'). This is called the bottleneck or reference case analysis. The latter term is used here because, in a later stage of the analysis, the effects of measures and strategies will be determined by comparing the results of that strategy with the reference case situation. The performance of the measure or strategy is assessed by using evaluation criteria related to the management objectives.

For the future (the planning horizons), the reference case comprises the present infrastructure, to which measures are added that have already been executed or decided, together with projected demands.The effects of autonomous developments in the SES are accounted for in these projected demands.

### 5.2.3. Identification and Screening of Measures

Once the base case and reference case have been defined, and problems and bottlenecks identified, measures are developed to alleviate the problems. In the identification of measures, a distinction is made between various categories of measure, as listed in Section 1.3.

An inventory of all possible kinds of action that can be taken will in general result in hundreds of measures. In most cases it will not be possible to analyse all of them in detail. A screening process is needed to select the most promising ones. This can be done in several ways. As mentioned in various chapters of this book, separate optimization models can be used to limit the solution space. It can also be done by using the computational framework developed for the project but limiting the analysis to a few criteria, such as benefit and costs. The measures with the highest benefit–cost ratio will be labelled as the most promising ones. A third kind of screening analysis is to apply a qualitative judgment, for example in terms of criteria effectiveness, efficiency, legitimacy and sustainability. Box E.6 explains these criteria.

The aim of the screening process is to get some feeling of how likely the measures are to alleviate the present and expected problems. No final judgment is given and no measures will be discarded in the process.

The screening of measures is a cyclic process. Assessing the measures will contribute to a better understanding of their effectiveness and new ones may be identified (comprehension loop). Combinations of measures may be considered for specific parts of the WRS, for instance for solving the water quality problems in a sub-basin. The result of the screening process is a set of promising measures that can be used for strategy design. The whole process of base case and reference case analysis and screening is depicted in Figure E.10.

**Box E.6. Criteria**

*Effectiveness.* Meaning that the measures to be taken are those which solve the most serious problems and have the highest impact on the objectives. Measures to prevent the problem will be preferred to those that solve it. Similarly, measures that solve the problem are preferred to those that only control it.

*Efficiency.* The measures to be taken should not meet the explicit objectives at the expense of other implicit objectives. The cost-benefit analysis (at the national level) is one indicator of efficiency. An example is to issue a law that forces industrial firms to incur the full cost of end-of-pipe treatment. In Egypt, this would improve the Nile-system water quality, and thus improve health and avoid environmental damage, but on the other hand it might impose high costs to the firms, possibly resulting in loss of employment. An efficient decision may be to opt only for cost sharing rather than full cost recovery.

*Legitimacy.* Measures to be included in the strategy should not rely on uncertain legal/institutional changes. Measures should also be as fair as possible, thus reducing public opposition so that they will be favoured by as many stakeholders as possible. For instance, reducing drainage water pollution through subsidizing non-chemical pesticides will be more legitimate than applying penalties on excessive use of chemical pesticides.

*Sustainability.* The measures to be taken are those that improve (or at least maintain) the present environmental and socio-economic conditions for future generations.

### 5.2.4. Finalization of the Computational Framework

During the screening of measures, adjustments in the computational framework may be necessary. Some components may need to be defined in more detail than was realized in the inception phase, while for others a simpler approach may be justified. Additional evaluation criteria may also be introduced to better discriminate the effects of different measures.

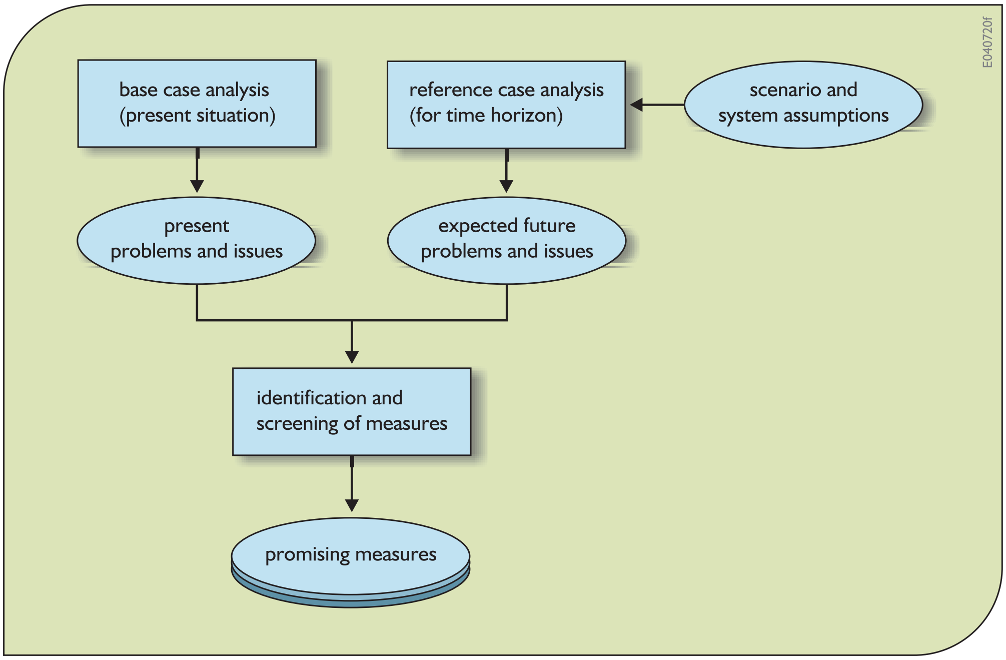



**Figure E.10.** Elements of the preliminary analysis.

# 6. Selection Phase

In the selection phase, promising measures are combined into strategies. The effects of various strategies are assessed and a limited set of promising ones is defined. For these promising strategies, the effects are assessed in more detail. The sensitivity of these effects to (uncertain) assumptions in the analysis is then assessed, and finally the results of the selected strategies are presented to the decision-makers. Figure E.11 highlights these steps, and the selection process is depicted in Figure E.12.

## 6.1. Strategy Design and Impact Assessment

Strategy design is the development of coherent combinations of promising measures to satisfy the management objectives as far as possible. As there are generally many criteria related to these objectives, and probably many expressed in different units, the strategy design is not a simple process. Relations among combinations of measures and their scores on the evaluation criteria are usually non-linear. As there are no general methods to compensate a lower score on one criterion with a higher score on another, the optimum itself is not uniquely defined. Sustainability should be considered in the assessment of measures and strategies.

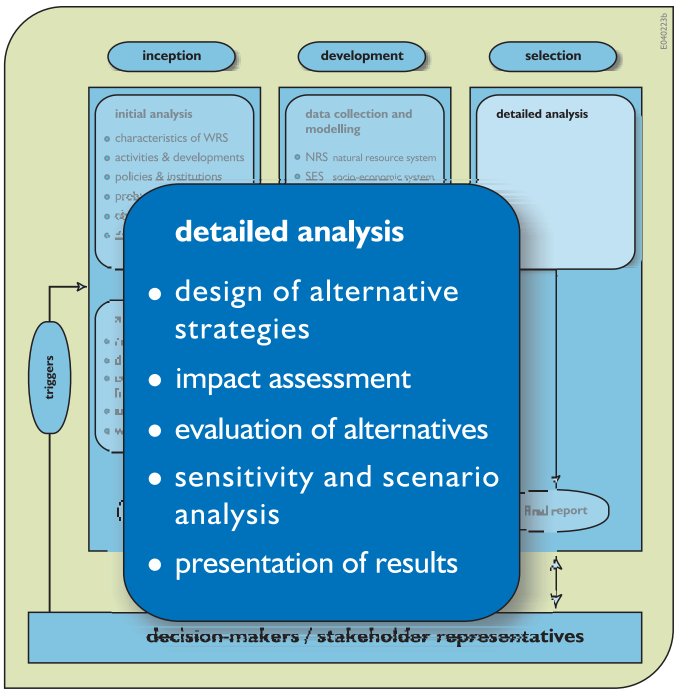



**Figure E.11.** Components of the selection phase.

The design of strategies is an iterative process. One could start by developing strategies on the basis of a single objective like, for example, food production or maximum net economic benefit. These strategies define

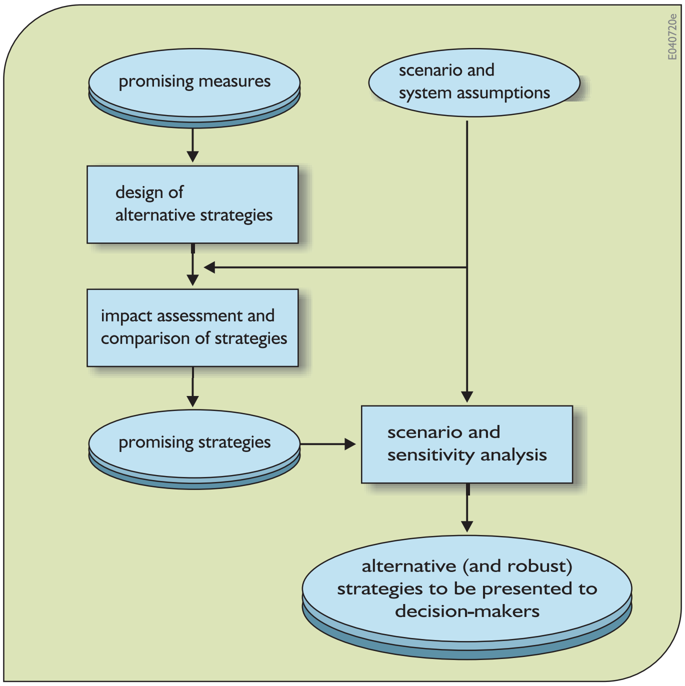



**Figure E.12.** Activities in the selection phase.

the boundaries of the solution space. Comparison of the impacts of these strategies can lead to the construction of compromise strategies by changing elements in the strategy. A resulting loss with respect to one criterion is then compared with gains to another.

The developed computational framework may be used for the assessment of selected strategies. Input data sets will be prepared for the computations, defining the modification of the water resources situation (strategy) and the assumed water demand (scenario). Such a data set will be referred to as a case. Each case is associated with a unique combination of a WRS, its demand situation (scenario) and its hydrometeorological boundary conditions.

## 6.2. Evaluation of Alternative Strategies

Strategies are compared on the basis of their scores on the evaluation criteria that were chosen to characterize the management objectives (see Section 4.1.3). To facilitate the comparison, the number of evaluation criteria should be limited. Criteria have to be comprehensive (sufficiently indicative of the degree to which the objective is met) and measurable, i.e. it should be possible to assign a value on a relevant measurement scale. Where possible, criteria should be aggregated; for example, some financial criteria might be processed into a single value when distribution issues are not important. It is usually impossible to express all criteria in a single measurement scale such as a monetary value. Criteria related to the quality of the environment can often be expressed only in non-monetary terms. This should however be done in such a way that a ranking is possible on the basis of the criteria.

Generally, there will not be a single strategy that is superior to all other ones with respect to all criteria used in the assessment. That means that an evaluation method is required for the ranking of alternative strategies (see Chapter 10). The evaluation can be an 'intuitive' one that gives implicit weights to various criteria and, depending on these weights, selects a preferred strategy.

In multi-criteria evaluation methods, formal decision rules are used to obtain a ranking. The weighted summation is one simple method. The best alternative is the one that gives the best weighted score.

## 6.3. Scenario and Sensitivity Analysis

Before drawing conclusions from the impact assessment, it is necessary to analyse the effects on the results of changes in the assumptions about the scenario variables and model parameters. The analysis of the influence of the exogenous variables is called scenario analysis; analysis of the influence of model parameter values (the system assumptions mentioned in Section 4.2.4) is called sensitivity analysis (Chapter 9).

Scenario analysis is done by repeating the impact assessment for the selected strategies under different scenario assumptions. If the selection of a different scenario would significantly change the attractiveness of the selected strategies, then additional study is required to reduce the uncertainties in the scenario assumptions. The sensitivity of the results for a change in model parameters and assumptions is done in a similar way.

## 6.4. Presentation of Results

Presentation of the selected promising strategies to decision-makers may be by means of briefings, presentations, summary reports or other media. The level of detail and the way results are presented have to be selected carefully. The presentation should give a good overview of the results, without an overdose of information or detail. Visual aids such as score cards and interactive computer presentations of study results are extremely helpful for a discussion of the results of the analysis.

The results of selected strategies can be presented in matrix form on 'scorecards'. The columns of the scorecard represent the alternative cases used in the analysis. The rows represent the impact of different alternatives with respect to a given criterion. An example is depicted in Figure E.13. Scorecards can contain numbers only, or the relative value of the criteria can be expressed by a colour or shading to obtain a visual picture of the relative order of the alternatives for various criteria. This can also help to detect clusters of criteria for which alternatives have a consistently better score. The presentation of the results in scorecards allows a decision-maker to give each impact the weight considered most appropriate.

Two examples of objectives and criteria are shown in the following two boxes. Box E.7 lists the objectives and criteria adopted in a water resources management study in West Java. Box E.8 is a scorecard used in Egypt for the development of the National Water Resources Plan. This example does not show alternative strategies; only the chosen strategy 'Facing the Challenge' has been given. Including the values of the base case (present situation) and reference case (future situation if no new measures are taken) provides a good overview of the new strategy's impacts.

## 7. Conclusions

This appendix has described the approach used by WL | Delft Hydraulics to assess water resources systems and develop plans and management strategies for them. Like many other such organizations, WL | Delft Hydraulics has been actively involved in numerous regional water resources planning and management projects throughout the world, mostly in developing countries. The approach described in this appendix illustrates how these projects are conducted, and the major factors that are considered while conducting them. The effects and impacts of some projects are only local and require consideration of only a few sectors of the economy. Other, more comprehensive projects have had national or international impacts. Clearly each water resources system is unique with respect to its management issues, problems and institutional environment.

| | promising strategies | | | |
|---|---|---|---|---|
| | agricultural strategy | industrial strategy | anti pollution strategy | mixed strategy |
| strategy components: | • irrigation<br>• water storage<br>• pumps<br>• etc. | • water storage<br>• groundwater use<br>• canal improvement<br>• etc. | • water conservation<br>• purification<br>• tax on water use<br>• etc. | • water storage<br>• purification<br>• etc. |
| impacts on criteria: | | | | |
| total investment costs m euro/yr | 300 | 400 | 700 | 700 |
| total benefits m euro/yr | 1200 | 700 | 100 | 1000 |
| increased agricultural production m ton/yr | 800 | 150 | 50 | 600 |
| drinking water price euro/m$^3$ | 1.40 | 0.90 | 1.20 | 1.10 |
| pollution ppm | 150 | 220 | 35 | 70 |
| power production MW | 200 | 1200 | 50 | 800 |
| fisheries ton/yr | 70 | 20 | 80 | 40 |
| safety from flooding % | 99 | 98 | 96 | 99 |

best middle worst

E040720b

**Figure E.13.** Example of a scorecard.

**Box E.7. Example 1: Objectives and criteria adopted in West Java WRM study**

| *OBJECTIVES* | *EVALUATION CRITERIA* |
|---|---|
| **SOCIO-ECONOMIC OBJECTIVES AND CRITERIA** | |
| **1. Improve employment** | Increase of employment by WRM strategies:<br>– number of permanent jobs (#)<br>– number of temp. jobs (mn/yr) |
| **2. Increase income of people**<br>– improve income position of farmers<br>– improve equity in income distribution. | <br>– farmer net income (Rp/yr)<br>– difference in benefits of WRM strategies per capita between:<br>kabupatens (%)<br>urban/rural areas (%)<br>income groups (%) |
| **3. Increase the non-oil export production** (shrimps, tea and rubber). | – export value (Rp/yr) |
| **4. Support economic development** in an economically efficient way. | – total annual. benefits (Rp/yr)<br>– total annualized costs (Rp/yr)<br>– B/C ratio (-)<br>- IRR (%)<br>– NPV (Rp/yr)<br>– total capital required (Rp)<br>– foreign currency required (%)<br>– total construction costs (Rp)<br>– total O&M costs (Rp)<br>– sectoral value added (Rp/yr)<br>– GRP (Rp/yr) |
| USER-RELATED (SECTORAL) OBJECTIVES AND CRITERIA | |
| **1. Increase agricultural production** (3% per year) | – padi (ton/yr)<br>– palawija (ton/yr)<br>– export value of crops (or import substitution) (Rp/yr)<br>– unit costs water supply (Rp/m$^3$)<br>– % failure meeting demand (%) |
| **2. Increase power production** | – installed capacity (MW)<br>– power production (GWh/yr)<br>– failure meeting firm power (%)<br>– price of power prod. (Rp/Kwh)<br>– energy production value (Rp) |
| **3. Increase fish production** | – fish produced (ton/yr)<br>– fish pond area (ha)<br>– export value (Rp/yr) |
| **4. Support industrial development**<br>• water supply for industry (full supply)<br>• provision of opportunity for discharge of waste water | – amount of supply (m$^3$/s)<br>– cost of water supply (Rp/yr)<br>– unit costs water supply (Rp/m$^3$)<br>– level of failure to meet demand (%)<br>– cost to maintain water quality standards (Rp/yr) |
| **5. Enhance water-related recreation** | |

*(Contd.)*

**Box E.7 Example 1: Continued**

| *OBJECTIVES* | *EVALUATION CRITERIA* |
|---|---|
| **ENVIRONMENTAL AND PUBLIC HEALTH RELATED OBJECTIVES AND CRITERIA** | |
| **1. Improve public health** | – supply (l/day/ capital) |
| • improve drinking water supply | – % of people connected (Rp/m$^3$) |
| – urban: BNA, IKK and major city programs: 60 l/cap/day, serving 70% | – price of drinking water (%)<br>– % failure meeting demand (%) |
| – rural: 55% | |
| • improve flushing (1 litre/s/ha in urban area) | – volume of flushing water (m$^3$/s)<br>– unit costs (Rp/m$^3$)<br>– % failure meeting demand (%) |
| **2. Improve/conserve natural resources and environment** | – area severely eroding (ha)<br>– erosion (mm/year) |
| • erosion and sedimentation control (erosion $<$ 1 mm/yr) | – sediment yield (tons/yr)<br>– reforested area (ha) |
| • conservation of nature<br>• water quality | – replanted area (ha)<br>– terraced area (ha)<br>– ratio of external wood supply to total wood demand (%)<br>– concentration water quality parameters (ppm) |
| **3. Provide flood protection** (return period: depending on value of endangered area) | – return period (years)<br>– flood alleviation benefits (reduced damage) (Rp/year)<br>– flood control cost (Rp/year)<br>– number of people in endangered areas (#)<br>– flooded area (ha) |
| **PLANNING AND IMPLEMENTATION RELATED OBJECTIVES AND CRITERIA** | |
| 1. Take care of maximum agreement with existing policies in other fields of planning (e.g. economic regional planning). | – deviations from/conflicts with existing policies |
| 2. Maximize flexibility of proposed strategy. | – degree to which strategy can be adjusted to changes in demands, standards, technological innovations |
| 3. Maximize reliability of proposed strategy. | – degree of certainty with which proposed strategy will meet the realization of objectives |
| 4. Provide sufficient acceptance of proposed strategy by public, interest groups and executing authorities. | – degree of acceptance by parties involved |
| 5. Take care of maximum agreement of proposed strategy with existing competence and responsibilities of agencies concerned. | – deviations from/conflicts with existing competence and responsibilities |

Note: GRP = Gross regional product
IRR = Internal rate of return
B/C = Benefit-cost ratio
NPV = Net present value
O&M = Operation and maintenance

○ kabupaten = Indonesian administrative unit ○ padi = rice crop ○ palwija = non-rice crop ○ Rp = Rupiah

### Box E.8. Example 2: Score-card for Egyptian National Water Resources Plan study

| | *Unit* | *1997 Base* | *2017 Reference Case* | *Strategy Facing the challenge* |
|---|---|---|---|---|
| **General (middle scenario)** | | | | |
| Population | Million | 59.3 | 83.1 | 83.1 |
| Urbanization | Ratio | 0.44 | 0.48 | 0.48 |
| GDP at economic growth of 6% | Billion LE | 246 | 789 | 789 |
| **Economic development objectives** | | | | |
| **Agriculture:** irrigation area | M feddan | 7.985 | 11.026 | 10.876 |
| Gross production value | Billion LE | 34.46 | 35.76 | 38.50 |
| Crop intensity | Ratio | 2.1 | 1.5 | 1.7 |
| Net value production per *feddan* | LE/feddan | 2,812 | 2,075 | 2,153 |
| Net value production per unit of water | LE/$m^3$ | 0.64 | 0.66 | 0.60 |
| Export/import value | Ratio | 0.09 | 0.12 | 0.20 |
| **Industry:** Costs of polluted intake water | LE/$m^3$ | 0.65–1.10 | 0.65–1.10 | 2.00 |
| Wastewater treatment costs | LE/$m^3$ | 0.22–0.50 | 0.22–0.50 | 1.00 |
| **Fishery:** production (index 100 in 1997) | index | 100 | 86 | 95 |
| **Tourism:** navigation bottlenecks | Index | 100 | 114 | 0 |
| **Social objectives** | | | | |
| **Create living space in desert areas** | % of total pop | 1.5% | 23% | 22% |
| **Employment and income** | | | | |
| Employment in agriculture | M pers.year | 5.01 | 6.24 | 7.30 |
| Employment in industry | M pers.year | 2.18 | 4.99 | 4.99 |
| Average income of farmers | LE/yr | 5,362 | 4,629 | 4,309 |
| **Drinking water supply** | | | | |
| Coverage | Percentage | 97.3% | 100% | 100% |
| **Sanitation** | | | | |
| Coverage | Percentage | 28% | 60% | 60% |
| **Equity** | | | | |
| Equity water distribution in agriculture | –, 0, + | 0 | + | + |
| **Self sufficiency in food:** cereals | Percentage | 73% | 53% | 46% |
| **Meeting water needs** | | | | |
| **Water resources development** | | | | |
| Available Nile water | BCM | 55.8 | 55.5 | 55.5 |
| Abstraction deep groundwater | BCM | 0.71 | 3.96 | 3.96 |
| **Water-use efficiency of Nile system** | | | | |
| Outflow to sinks from Nile system | BCM | 16.3 | 17.6 | 12.5 |
| Overall water-use efficiency of Nile system | Percentage | 70% | 67% | 77% |
| **Water in agriculture** | | | | |
| Supply-demand ratio (1997 assumed 1.0) | Ratio | 1.00 | 0.80 | 0.92 |
| Water-availability per *feddan* in Nile system | $m^3$/feddan/yr | 4,495 | 3,285 | 3,866 |
| **Public water supply** | | | | |
| UFW losses | Percentage | 34% | 34% | 25% |
| Supply/demand ratio | ratio | 0.67 | 0.76 | 1.00 |
| **Health and environment** | | | | |
| **Pollution and health** | | | | |
| E-coli standard violation (1997 = 100) | Index | 100 | 121 | 110 |
| Water quality of shallow groundwater | –, 0, + | 0 | – | – |
| **Ecology and sustainability** | | | | |
| Sustainability: use of deep groundwater | abstr/pot | 0.15 | 1.00 | 1.00 |
| Condition of Bardawil (Ramsar site) | –, 0, + | + | – | + |
| Condition of Coastal Lakes | –, 0, + | 0 | – | 0 |

Note: UFW = Unaccounted for water (the water that is lost in the system)

○ feddan = 0,42 ha ○ LE = Egyptian pound

Project planning and analysis approaches must adapt to these situations. Hence, each project will differ, and will no doubt need to deviate from the general approach described in this appendix. Other approaches are possible and may be equally effective. What remains important in all cases is the establishment of a comprehensive, systematic process of analysis together with constant communication among planners, decision-makers and the interested and affected public. The end result should be an improved, sustainable and equitable water resources development plan and management policy, appropriate for the region and its people.

# Index

**Notes:** text in boxes is indicated by **bold page numbers** and in figures and tables by *Italic* page numbers.


**A**

Adaptive management 24, 257, 322, 421, 466, 504, 559, 576
Advective transport 385, 522
Aeration 436
Agriculture 11, 16, *51*, 113, 325, 405, 458, 475, 501, 564, 581, 675, 608
Algae
  Eutrophication 3, 15, 393, 450–59, 474–78 507–520
  Modelling 30, 59, 412, 497–514
Allocations 20, 60, 83, *85*, 138, 156, 236, 303, 356, *364*, 389
Alternatives 23, 45, 52, 59, 214, 284, *286*, 295–96, *313*, 359–60, 613
Anaerobic 393, 452, 497
Annual value 82
Aquatic systems 386, 464, 501, **656**
Aquifer 6, 50, *51*, 199, *333*, 542, *656*
Artificial neural networks 41, 147, 148
Autocorrelations 198, 201
Autoregressive-moving average (ARMA) models 213
Average values 170, 283

**B**

Benefits 3, 46, 60, 81, *85*, 293, 342, 513, **597**, 627, *629*
Benefit–cost
  Analyses 298, 360, 627
  Ratio **674**, 669
Bias 150, 176
Biochemical oxygen demand (BOD) 390, *449*, *451*, 478
Boundary conditions 74, 260, 536
Box-Jenkins models 213
Budget 302, 368, 525, 578

**C**

Calibration 69, 118, **120**, 148, *257*, 260, 441
Capacity expansion 61, 98, 368
Capital 16, 82, 454
Capital recovery factor 83
Case Studies 4–8, 467–478
Censored data 193–95
Central limit theorem 276
Chance constraints 235
Climate change 4, *257*, 542, 588–90, 608
Coasts
  Beaches 538
  Management 540
  zones 535
Coefficient of skewness 174, 176, 178, *190*
Coefficient of variation 174, *178*, 189
Coliform
  Bacteria 378, *391*, 435
  Count 379
Collective goods 300
Compensation criterion 300
Competitive conditions 299
Concave functions 84, 89, 130
Concentration gradient 385
Confidence intervals 184, 219, 260
Constraint method 311–20
Constraints 39, 60, 64, 87, 266, 296
Constituents
  Conservative 390
  First-Order 390
Continuous variables 70
Convex functions 86, 130
Correlation 232, 275, 276, 450, 568
Cost-effectiveness 298, 360, 627
Cost 10, 53, 62, 82, *98*, *123*, *130*, 281, 299–*301*
Covariance 206, 271
Critical value 169, 401
Crop 581, *583*, 592
Cross-correlation 196

**D**

Data
  Collection *560*, 564, 570
  Management 565
  Mining 163
Dead storage 342, *343*, *364*, **366**
Decision-making 22, 25, 29, 47, 92, 265, 294–95, 462–64
Decision
  Support systems 26, 47–*49*, 78
  Uncertainty 24, 72, *258*, 260
  Variables 60, 81
Deficit value 315
Degrees of freedom 76
Demand
  Agricultural 357
  Function 358
  Industrial 356–57
  Municipal 356
Depression storage 439–*440*
Depth 102, 154, 173, 261, *268*, 306, 402, 491
Design 255, 272, 388, 437–*39*, 456, *493*, *509*, 560, 565
Deterministic equivalent 236
Deterministic modelling 231
Differential equations 89, 416

Differential persistence 209
Dimensionality 97
Disaggregation 209, 211
Discount factors 82
Dispersion
  Bulk 399
  Dispersive transport 385, 522
Dissolved oxygen 41, 390–93, 435
Distribution systems 427
Diversions 83, *234*, 239, 355, *364*
Dominance 310–11, 585
Droughts
  El Niño and La Niña 587
  Indices 590–95
  Management 581, 596
  Triggers 596
  Virtual exercises 596
Dynamic expansion 368
Dynamic Programming
  Backward-moving 92
  Deterministic 240–41
  Dimensionality 97
  Forward-moving 95–96
  Network 90
  Stochastic 133

**E**
Ecology
  Ecological criteria 306–308
Economics
  Criteria 298
  Long-run benefits 303
  Loss functions 305
  Short-run benefits 303
Economies of scale 98
Ecosystems 3, 21, **26**, 306, 377, 500
Effective 26, 71, 331, 381, 559
Environmental
  Criteria 305–306
  Impacts 303
Epilimnion 407
Equity 28, 142, 309, 388, **675**
Equivalent 82–83, 284, 343
Error 175, 176, **257**, 260, 276, 277
Estimation 179–82, 299–300, 339, 348
Estimators 173–76
Eutrophication 252, 287, 393, 450–59, 474–78
Evaporation 326, 332, 346–52, *355*
Excess value 127
Expansion 16, 368
Expectations 4, 173, 559
Expected value 169, 173, 248, 321

**F**
Failure 45, 217, *218*, 348, 631–35, **674**
Feasible 63, 240, 246–*47*
Finite difference 416
Firm yield 347
First-order 53, 198, 238, 270–71
Fixed costs 98, 130
Flood
  Damage 342, *361*
  Flood storage capacity 359, *343*
  Flood dikes and levees *607*, *632*
  Management 51
  Protection 43, 360, *362*
  Reservoir 169, *170*, *216*
  Return period 197, 362, *366*
  Risk 170, 359
  Storage 326
Flow augmentation 295, 296
Flux 386, 402, 513
Fractional factorial design 272
Fuzzy
  Models 142
  Optimization 135
  Sets 138, *140*

**G**
Generalized likelihood estimates 181, 192
Genetic algorithms 147, *156*, *158*, 341
Genetic programming 148, 159–63
Global optimum 89, 158
Goal attainment 314–15
Goal programming 315
Goodness-of-fit 176, 186, 277
Groundwater
  Flows 326, *334–35*, 509
  Interactions 336, 513
  Management 540, 542
  Modelling 336
Growth 9, 16, 22, 195, 390, 414, 468

**H**
Hazen plotting position 183
Heat 389, 390
Heavy metals 398–99, 452
Hill-climbing methods 84
Hydroelectric power 18, 357–59
Hydrological
  Models 328–29
  Parameters 329
  Processes 329
    Groundwater 333
    Surface water 329, 331
Hypolimnion 407
Hypothesis 282, 446

**I**
Impervious 380, 440, 496
Incremental flow 84, 419
Index-flood method 195, 196
Indifference analysis 284, 313
Inequality constraints 89
Infiltration 328, 329, 437, 441, 496
Inflation 70, 198
Interest 82–83
Internal rate of return **674**
Interquartile range 173
Investments 27, 28, 59, 83, 660
Irreversibility 284
Irrigation 3, 4, 70, 211, 298, 357, 474, 648

**J**
Joules 359

**K**
Kilowatts 359
Knowledge uncertainty 258, 260, 420
Kolmogorov-Smirnov 183

**L**
Lagrange multipliers 71, 83–86
Lake
  Outflow 406
  Quality 406
  Recreation 362
Land 59, 204, 300, 377, 590
Latin hypercube sampling 275, 278
Level 20, 48, 103, 169, 380, 501, 542, 650
Lexicography 313
Light 32, 400, 413, 430, 571
Linear programming
  Piecewise linearization 126, 128
Linearizations 127
Loading 10, 41, 380, 445, 446
Local optima 89, 162
Log-likelihood 180
Long-run 23, *303*, 305, 368
Loss 6, 21, 272, 341, 377, 534, 584

**M**
Management
  Adaptive 24, 31
  Approaches 24
  Models 28, 30
Marginal 82, 172, 208, 299
Market prices 299, 300–301
Markov
  Chains 198–200
  Processes 198–200
Mass diagram 343, *344*
Mass transport 385
Maximum likelihood estimates 186, 188
Mean 103, 169, 174, 265, 271, 320, 351, 393
Median 169, 173, 282, 590
Method of moments 180, 181
Michaelis-Menten kinetics 393, 395
Mineralization 408, 498
Model
  Adequacy 182–84
  Calibration 118–120, **121**
  Data 465
  Development 45, 46, 71
  Dynamic 153
  Errors 259, 260
  Fuzzy 138
  Journal 72
  Linear 70, 113
  Output 255, 258, 283
  Optimization 64–65, 83
  Selection criteria 380–81
  Shared-vision 31
  Simulation 64
  Steady-state 386–88
  Structure 255, 260
  Success 466–67
  Verification 41, 72, 155
Moments
  L-moments 176–78

Monitoring
  Network 564, 565
  Plan 563
  Sampling 564, 566
Monte Carlo
  Sampling 273, 274, 275
  Simulation 233–35, *274*
Multi-criteria analyses
  Constraint method 311–13
  Dominance 310
  Goal-attainment 314
  Goal-programming 315
  Indifference analysis 313
  Interactive methods 315
  Lexicography 313
  Satisficing 313
  Weighting method 311
Multiple criteria or objectives 321
Multiple purposes 20, 231, 355, 461, 624
Multiple yields 347, 352
Multiplier 88
Multivariate models 209

**N**

National economic development 309
Navigation 21, 163, 164, 356, 613, 619, 623
Network 90, 148, 432, 434, 565
Nitrogen
  Cycle *409*
  Models 381, *409*
Non-inferior 310, 312
Non-point
  Models 296
  Quality 296
  Runoff 356, 427
Normal distributions 275
Nutrients
  Cycling 408
  Models 412–13

**O**

Objective 43, 59, 63, 138, 156, 179, 256, 293, 364, 605, **656**
OMR Costs 83
Operating policy 59, 60, 107, 214, 216, 248
Operating rule 352
Operational hydrology 225
Operations Research 56, 133
Opportunity cost 27, 302
Optimal 24, 54, 64, 81, 161, 248, 264, 310
Optimization 59, 60, 81, 135, 239, 309, 368, 453
Organics
  Micropollutants 397
Over-year 203, 348
Oxygen 41, 297, 378, 390, 435, 450, 514, 562

**P**

Parameter
  Calibration 120, 150, 260, 263, 333
  Uncertainty 260
  Value 260
performance measures 59, 103, 135, 281, 352
Peak flow 360, 362, 486, 628, 630
Performance criteria 20, 62, 293, 295
Phosphorus
  Cycle 410, *411*
  Model 410
Photosynthesis 392, 415, 452, 490
Plan
  Formulation 61–63
  Implementation 32, 75, 164, 467
  Selection 61, 63–64
Planning
  Approaches 24
  Economic 27–28
  Financial 27
  Integrated 26, 31
Plant capacity 358
Plant factor 358
Plotting positions 183, 186
Policy 5, 19, 26, 86, 107, 164, 242, 311, 322, 379, 421, 428, 598, **647**
Population 3, 6, 175, 204, 319, 405, 517, 531, 574, 598, 652
Precipitation 70, 205, 257, 326, 399, 590
Present value 63, 82, 298, 369, **674**
Price 86, 220, 300, 302, 428, 664
Primary treatment 435
Principal of optimality 97
Probability
  Conditional
  Distribution 172
    Beta 182
    Cumulative 171
    Density 171
    Exponential 184
    Functions 179
    Gamma 187–89
    Gumbel and GEV 190
    Lognormal 186–87
    Log-Pearson Type 3 189–90
    Normal 186–87
  Exceedance 183
  Joint 171
  Marginal 172, 208
  Transition 199
  Unconditional 200
Production 3, 7, 18, 86, 298, 309, 357, 386, 434, **673**
Project
  Management 72
  Scheduling 368
Public 5, 12, 43, 83, 266, 295, 456, 463, 546, 596, 609, 645

**Q**

Quality
  Impacts 448
Quantity
  Flows 271, 325
  Storage 27, 327
  Velocities 325
Quantiles 173, 176, 278

**R**

Radioactive substances 396, 400
Radionuclides 400
Rainfall
  Design rainfall 438–39
  Distribution in space and time 438
  Models 444–45
  Synthetic rainfall 438
Random variables
  Independent 172, 180
Rank 283, 313, 348
Rate 81, 82, 83, 153, 170, 306, 331, 380, 432, 491, 585, 617
Reaction 7, 386, 393, 415
Re-aeration 384, 392, 415
Recreation 325, 362, 516
Recurrence 348, 585, 618
Recursive 90, 94, 97, 207
Redundant constraints 89
Regional 8, 44, 59, 176, 255, 298, 427, 462, 517, 568, 585, 619
Regionalization 176, 195
Regret 31
Reliability 45, 294, 321
Reservoir
  Active storage volume 103
  Capacity 115
  Dead storage volume 353
  Flood storage volume 353
  Operation 102
  Optimization *68*
  Over-year storage 203
  Release rule 102
  Simulation *68*, 352
  Yield 114
  Within-year storage 353
Resilience 111, 321, 535, 559
Respiration 392, 393, 415, 498
Response 4, 77, 153, 268, 302, 368, 380, 444, 490, 559, 585, 620
Return period 197, 362, 438, 493, **674**
Revenue 28, 87
Risk 359, 451, 609, 627, 631, 634
Rivers 483–84
River basins
  Management 328
  Simulation 215
Root zone 332, 513
Routing 341, 342, 440, 627, 662
Rule curves 102, 248, 353
Runoff 439–441, 444

**S**

Safe yield 347
Sampling
  Monte Carlo 273–75
  Uncertainty 275–76
Satisficing 313
Scale
  Processes 75, *76*
  Spatial 22, *327*
  Temporal 23, *327*
Scarcity 20, 298, 505
Scheduling 368–370, 455
Screening 30, 251, 359, 562, 569, 669
Search 7, 63, 97, 147, 192, 319, 454, 461
Second-order conditions 180

Sediment
Bed load 403
Burial 402
Resuspension and Sedimentation 401
Settling 396, 397, 398
Suspended load 492
Seepage 103, 257, 336, 504
Sensitivity
Analyses 261
Coefficients 267
First-order analysis 270
Sequent peak analyses 343
Settling
Detritus 409
Inorganic 409
Sediment 398
Sewer systems 448, 478
Silica Cycle 411
Simulation
Accuracy 416
Methods 416
Uncertainty 420
Social criteria 24, 308
Stakeholder 4, 45, 295, 364, 463, 535, 561, 626
State variable 90, 97
Static analyses 327
Stationarity 197
Statistical 41, 69, 147, 183, 231, 259, 320, 329, 380, 438, 465, 564
Steady-state 107, 200, 353, 386
Step method *73*, 315
Stochastic
Model 216
Optimization 239
Processes 197
Simulation 214
Variables 214
Storage zones
Active 343
Dead 362
Flood 360
Over-year 353
Stratified 407
Within-year 353
Storage-yield
Functions 344, *346*
Storms
Stormwater 447
Stratification 407
Streamflow
Estimation 339–40
Gauge 328, *340*
Generation 203
Routing 341
Synthetic values 203
Streeter and Phelps 390
Student's *t*-distribution 222
Substitution 173, 210, 298
Sustainability 20, 23, 322, 483, 503
Synthetic Streamflow 203, 205
Systems analysis 12, 30, 476
Systems Approach 461

**T**

Target 104, 278, 575
Temperature 76, 389–90
Thermocline 407
Thomas-fiering model 213
Time periods 327
Time series 197, 259, 437
Time value 82
Toxic chemicals 396, 534
Tradeoffs 59, 280, 293, 296–97, 355, 388
Transition probability 199, 238
Transpiration 326, 332, 439, 513
Transport 14, 118, 162, 293, 367, 385, 448, 464, 491, 626
Trial and error 63, 103, 157, 214, 421, 454

**U**

Uncertainty 254, *258*
Analyses 261
Model 262
Parameter 260, 262
Uniform 83, 157, 214, 232, 277, 333, 434, 523, 612
Urban
Drainage 437
Model 441, 442
Pollutant loading 445
Runoff 439

**V**

Validation 72, 151, 570
Variability 258, 259
Variables
Deterministic 169
Random 170, *174*, 214
Variance 174, 175, 232, 255, 320, 588
Verification 16, 41, 69, 154, 384
Vulnerability 111, 220, 319, 321, 559, 598

**W**

Wastewater
Collection 427, 434
Discharge 434
Treatment 435–*36*
Water
Groundwater 124, 333, 336, 513, 542
Polluted 21
Quality 434, 497, 506, *122*
Quantity 567
Surface water 331, 592
Watershed 23, 76, 103, 148, 193, 261, 297, 328, 379, 427, 466, 585, 608
Weibull plotting position 183, 185
Weighting method 311
Wetlands 510
Wilcoxon test 283
Willingness to pay 82, 299
Wilson-Hilferty transformation 189
Withdrawals 3, 355
Within-year
Periods 347, 354
Storage 347

**Y**

Yield 113, 114, 344, 347

**Z**

Zooplankton 410, 569

Droughts, floods and pollution are frequently viewed as constraints to economic and social development. How too little, too much or over-polluted water is managed can determine the extent to which this critical resource contributes to human welfare. A variety of management tools and approaches exist, however, that can help water and watershed managers identify development plans and management policies that best meet society's economic, environmental and social goals.

Over the past several decades both authors have been applying these tools to water resource system planning, development and management projects worldwide.

This book is based on experiences teaching at universities and in using these tools in practice. It uses case studies to introduce both the methods and demonstrates their practical application. It will serve as a valuable reference for both consultants involved in planning and development projects and as a text, with exercises, for university students.

About the authors:

Daniel P. Loucks, Professor of Civil and Environmental Engineering at Cornell University in the United States, has been actively involved in both the development and application of water resources systems models for over three decades. As a consultant to private, governmental and international organizations he has participated in regional water resources planning, development and management projects in Africa, Asia, Australia, Eastern and Western Europe, and North and South America.

Eelco van Beek, water resources specialist and manager of the freshwater systems group within WL / Delft Hydraulics, The Netherlands, has been involved in river basin planning and management projects all over the world, together with model software development and use, for over three decades. As a professor he headed the integrated water resources management program in the Faculty of Civil Engineering and Geosciences at the Technical University in Delft.

WL | delft hydraulics

UNESCO
United Nations
Educational, Scientific and
Cultural Organization

www.unesco.org/publishing

ISBN 92-3-103998-9

9 789231 039980