Daniel P. Loucks
Eelco van Beek

Water Resource Systems Planning and Management

An Introduction to Methods, Models, and Applications

Water Resource Systems Planning and Management

Daniel P. Loucks · Eelco van Beek

Water Resource Systems Planning and Management

An Introduction to Methods, Models, and Applications

With Contributions by
Jery R. Stedinger · Jos P.M. Dijkman · Monique T. Villars

OPEN Deltares UNESCO-IHE Institute for Water Education Springer

Daniel P. Loucks
Cornell University
Ithaca, NY
USA

Eelco van Beek
Deltares
Delft
The Netherlands

Foreword

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 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 threefold—from 1.8 to 6 billion people—water withdrawal during the same period increased sixfold! 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 % of the total. This is indeed very limited and on average is roughly equivalent to 42,000 km^3 per year.

If one looks at the past 30 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 m^3 per person per year, it has now dropped to 6000 m^3; 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 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 substantially reduce the number of human beings who have no access to safe drinking water and adequate sanitation facilities respectively. 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 analyzing 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 Profs. 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
Past Deputy Assistant Director-General, UNESCO
and Past Secretary, International Hydrological Programme
Past Rector Magnificus, UNESCO-IHE, Delft, The Netherlands

Preface

Water resource systems planning and management issues are rarely simple. Demands for reliable supplies of clean water to satisfy the energy, food, and industrial demands of an increasing population and to maintain viable natural ecosystems are growing. This is happening at the same time changes in our climate are increasing the risks of having to deal with too little or too much water in many river basins, watersheds, and urban areas. Societies are becoming increasingly aware of the importance of water and its management and use; their governing institutions are becoming increasingly involved in water resources development and management decision-making processes. To gain a better understanding of the complex interactions among all the hydrologic, ecologic, economic, engineering and social components of water resource systems, analyses based on systems perspectives are useful. While analyses of such complex systems can be challenging, integrated systems approaches are fundamental for identifying and evaluating options for improving system performance and security for the benefit of all of us.

Just how well we are able to plan and manage our water availability, quality, and variability is a major determinant of the survival of species, the functioning and resilience of ecosystems, the strength of economies, and the vitality of societies. To aid in the analysis of planning and managing options, a variety of modelling approaches have been developed. This book introduces the science and art of developing and applying various modelling approaches 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 issues and problems. Their purpose is to provide relevant, objective, timely and meaningful information to those who are responsible for deciding how we develop, manage, and use our water resources.

Readers of this book are not likely to learn the art of systems modelling and analyses unless they actually do it. The modelling approaches, examples and case studies contained in this book, together with the exercises offered at the end of most chapters, we believe and hope, will facilitate the process of becoming a skilled water resources systems modeler, analyst and planner. This has been our profession, indeed our hobby and source of enjoyment, and we can highly recommend it to others.

Water resource systems planning and management is a multidisciplinary activity. The modelling and analysis of water resources systems involves

inputs from the applicable natural and social sciences and from the people, the stakeholders, who will be impacted. It is a challenge.

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. Rather it is intended to introduce readers to the art of developing and using models and modelling approaches applied to the planning and managing of water resources systems. We have tried to organize our discussion 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 approaches, the choice of a particular model and solution method should be based on the knowledge of various modelling approaches and their advantages and limitations. There is no one best approach for analyzing all the issues one might face in this profession.

This book assumes readers have had some mathematical training in algebra, calculus, geometry and the use of vectors and matrices. Readers will also benefit from some background in probability and statistics and some exposure to micro-economic theory and welfare economics. 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 the Appendices (available on the internet along with the book itself). 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 reviews some examples of water resources systems projects in which modelling has had a critical role. These projects also serve to identify some of the current issues facing water managers in different parts of the world. Chapter 2 introduces the general modelling approach and the role of models in water resources planning and management activities. Chapter 3 begins the discussion of optimization and simulation modelling and how they are applied and used in practice. Chapter 4 focuses on the development and 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 analyzed and improved using simulation methods. The advantages and limitations of different

optimization/simulation approaches are illustrated using some simple water allocation, reservoir operation and water quality management problems.

Chapter 5 extends this discussion of optimization to problems characterized by more qualitative objectives and/or constraints. In addition, it introduces some of the more recently developed methods of statistical modelling, including artificial neural networks and evolutionary search methods including genetic algorithms and genetic programming. This chapter expects interested readers desiring more detail will refer to other books and papers, many of which are solely devoted to just these topics. Chapters 6 through 8 are devoted to probabilistic models, uncertainty and sensitivity analyses. These methods are useful not only for identifying more realistic, reliable, and robust infrastructure designs and operating policies for the 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 the information useful to decision-makers who must decide what to do given these tradeoffs among conflicting performance criteria that stakeholders care about. Chapter 9 on multi-objective modelling identifies various types of economic, environmental and physical objectives, and some commonly used ways of including multiple objectives in optimization and simulation models.

Chapter 10 is devoted to various approaches for modelling water quality in surface water bodies. Chapter 11 focuses on modelling approaches for multiple purpose water quantity planning and management in river basins. Chapter 12 zooms into urban areas and presents some ways of analyzing urban water systems. Finally, Chap. 13 describes how projects involving the analyses of water resource systems can be planned and executed.

Following these thirteen chapters are four appendices. They are not contained in the book but are available on the internet where this book can be downloaded. 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, and (D) flood management.

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 might include Chaps. 1 through 5, and possibly Chaps. 9 and 10 or 11 or 12 or 13 depending on the background and interest of the participants in the class. A second course could include Chaps. 6 through 8 and/or any combination of Chaps. 10 through 12, as desired. Exercises are offered at the end of each chapter, and instructors using this text in their academic courses can contact the authors for the solutions of those exercises if desired.

Many have helped us prepare this book. Jery Stedinger contributed to Chaps. 6, 7 and 8, Nicki Villars helped substantially with Chap. 10, and Jozef Dijkman contributed a major portion related to flood management. Tjitte Nauta, Laura Basco Carrera and Thijs Stoffelen contributed to Chap. 13. Others who offered advice and who helped review earlier chapter drafts include Vladan Babovic, Martin Baptist, Henk van den Boogaard, Herman Breusers, Harm Duel, Herman Gerritsen, Peter Gijsbers, Jos van Gils, Simon Groot, Karel Heynert, Joost Icke, Hans Los, Marcel Marchand, Tony Minns, Erik Mosselman, Arthur Mynett, Roland Price, Erik Ruijgh, Johannes Smits, Mindert de Vries and Micha Werner. Engelbert Vennix and Hans van Bergem created most of the figures and tables in this book. We again thank Deltares and all these individuals and others who provided assistance and support on various aspects during the entire time in 2005 and when this second edition was being prepared.

We have also benefited from the comments of Profs. 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, MA, Tineke Ruijgh-van der Ploeg at TU-Delft, Robert Traver at Villanova University in Philadelphia, and Jinwen Wang at Huazhong University of Science and Technology in Wuhan, all of whom have used earlier drafts of this book in their classes. Finally we acknowledge with thanks the critical support of Andras Szöllö-si-Nagy, recently retired as rector of UNESCO-IHE in Delft, NL, and the publishing staff at UNESCO for publishing the first edition of this book and making its electronic version free and unrestricted. We are especially grateful to Michael Luby of Springer for all his assistance and guidance, and to Deltares and UNESCO-IHE in Delft, NL, for sharing the cost of publishing this second edition. We have written this book for an international audience, and hence we are especially grateful and pleased to have this connection to and support from Deltares and UNESCO-IHE.

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.

Ithaca, NY, USA   Daniel P. Loucks
Delft, The Netherlands   Eelco van Beek
July 2016

Contents

Water Resources Planning and Management: An Overview

1.1 Introduction

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 of the world they are not able to meet even basic drinking water and sanitation needs. Nor can many of these water resource systems support and maintain resilient biodiverse ecosystems. Typical causes include inappropriate, inadequate and/or degraded infrastructure, excessive withdrawals of river flows, pollution from industrial and agricultural activities, eutrophication resulting from nutrient loadings, salinization from irrigation return flows, infestations of exotic plant and animals, excessive fish harvesting, flood plain and habitat alteration from development activities, and changes in water and sediment flow regimes. The inability of water resource systems to meet the diverse needs for water often reflect failures in planning, management, and decision-making—and at levels broader than water. Planning, developing, and managing water resources to ensure adequate, inexpensive, and sustainable supplies and qualities of water for both humans and natural ecosystems can only succeed if we recognize and address the causal socioeconomic factors, such as inadequate education, corruption, population pressures, and poverty.

Over the centuries, surface and ground waters have been a source of water supply for agricultural, municipal, and industrial consumers. Rivers have provided hydroelectric energy and inexpensive ways of transporting bulk cargo. They have provided people water-based recreational opportunities and have been a source of water for wildlife and their habitats. 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 crop production and relatively flat lands for the siting of roads and 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 cultural, 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 damages such as when the river is experiencing periods of droughts, floods, and heavy pollution. These costs and damages are physical, economic, environmental, and social. They result because of a mismatch between what humans expect or demand, and what nature offers or supplies. Human activities tend to be based on the "usual or normal" range of river flow conditions. Rare or "extreme" flow 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 extreme flow conditions will incur losses.

The planning of human activities involving rivers and their floodplains must consider certain hydrologic facts. One of these facts is that surface water flows and aquifer storage volumes vary over space and time. They are also finite. There are limits to the amounts of water that can be withdrawn from them. There are also limits to the amounts of pollutants that can be discharged into them. 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 users of the resource.

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

River basin, estuarine, and coastal zone managers—those responsible for managing the resources in those areas—are expected to manage those resources 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 hydrologic and hydraulic processes become more pronounced, and as stakeholder expectations of system performance increase in complexity. The focus or goal is no longer simply to maximize economic net benefits while making sure the distribution of those benefits is equitable. There are also environmental and ecological goals to consider. Rarely are management questions one-dimensional, such as how can we provide, at acceptable costs, more high-quality water to municipalities, industry, or to irrigation areas in the basin. Now added to that question is how would those withdrawals affect the downstream hydrologic water quantity and quality regimes, and in turn the riparian and aquatic ecosystems.

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 and manage more effectively.

This chapter reviews some of the issues requiring water resources planning and management. It provides some context and motivation for the following chapters that outline 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 this chapter.

1.2   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.

1.2.1   Kurds Seek Land, Turks Want Water

The Tigris and Euphrates Rivers (Fig. 1.1) created the "Fertile Crescent" where some of the first civilizations emerged. Today their waters are critical resources, politically as well as geographically. In one of the world's largest public works undertakings, Turkey's Southeast Anatolia Project includes 13 irrigation and hydropower schemes, and the construction of 22 dams and 19 hydroelectric power plants on both the Tigris and the Euphrates. Upon completion, it is expected to provide up to 25% of the country's electricity.

Its centerpiece, the Ataturk Dam (Fig. 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 (Fig. 1.3).

When the multireservoir 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 the 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 in anti-Turkish causes.

The region of Turkey where Kurd's predominate is more or less the same region covered by the Southeast 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 resource 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 50-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

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

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

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

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.

As everyone knows, the Middle East is currently (2016) witnessing considerable turmoil so who knows the fate of any water resources project in this region, including the one just described in Turkey and the following example in Jordan. One can only hope that the management and use of this scarce resource will lead to more peaceful resolutions of conflicts not only involving water but of other political issues as well.

1.2.2   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 (Fig. 1.4) are placing heavy demands on its scarce freshwater resources. This largely arid region receives less than 250 mm of rainfall each year, yet total water use for agricultural and economic activities has been steadily increasing. This plus 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. From the 1300 million cubic meters (mcm) 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 million cubic meters (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 from 200–300 mcm. Both in quantity and quality this water is unsuitable for irrigation and does not sufficiently supply natural systems either. The salinity of the Jordan River reaches up to 2000 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.

One result of this increased pressure on freshwater resources is the deterioration of the region's wetlands. These wetlands are important for water purification and flood and erosion control. As agricultural activities expand, 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 (White et al. 1999). Without these natural ecosystems, it will be difficult and expensive to sustain high-quality water supplies. The role of ecosystems in sustaining water supplies has largely been overlooked in the context of the region's water supplies. Vegetation controls storm water runoff and filters polluted water, and it 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 habitat for many plants and animals.

The Jordan River Basin just like most river basins should be evaluated and managed as a whole system, 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. However before such measures are required, one should assess the impact on local aquifer recharge, storage, and withdrawals.

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?
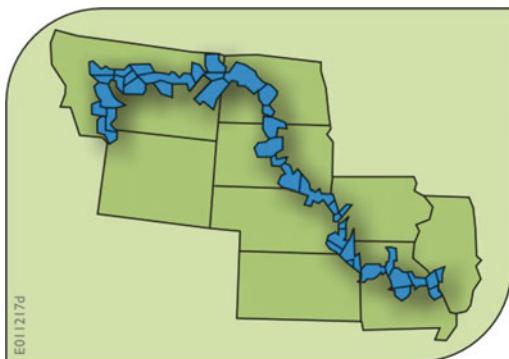
1.2.3 Mending the "Mighty and Muddy" Missouri

Nearly two centuries after an epic expedition through the Western US in search of a northwest river passage to the Pacific Ocean, there is little enchantment left to the Missouri River. Shown in Figs. 1.5 and 1.6, it has been dammed, diked, and dredged since the 1930s mainly 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 muddy. The conservation group American Rivers perennially lists the Missouri among the USA's 10 most endangered rivers.

Fig. 1.5 Major river basins in the continental US

Fig. 1.6 The Missouri Basin's Reservoirs (not to scale) constructed for navigation and flood control

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

Its lower one-third is a narrow canal sometimes called "The Ditch" that is deep enough for commercial towboats. Some of the river's banks are armored 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, and in some cases even survive.

Changes to restore at least some of the Missouri to a more natural state are being implemented. These changes add protection of fish and wildlife habitat to the list of objectives to be achieved by the government agencies managing the Missouri. The needs of wildlife are now 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 28 Native American tribes, can reach a compromise with the traditional downstream uses of the river.

1.2.4 The Endangered Salmon

Greater Seattle in the northwestern US state of Washington may be best known around the world for Microsoft, but residents know it for something less flashy: its dwindling stock of wild salmon. 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 could slow land development in one of the fastest growing regions of the U.S.

Before the Columbia River and its tributaries in NW US were blocked with dozens of dams, about 10–16 million salmon made the annual run back up to their spawning grounds (Fig. 1.7). In 1996, a little less than 1 million did. But the economy of the NW depends on the dams and locks that have been built in the Columbia that provide cheap hydropower production and navigation.

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

For a long time, engineers tried to modify the system so that fish passage would be possible. As shown in Fig. 1.8b, this included even the use of trucks to transport captured juvenile salmon around dams for release downstream. (It is not clear that the trucks will be there when the fish return to spawn upstream of the dams.) These measures have 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 trucked around the turbines. The costs of keeping these salmon alive, if not completely happy, are enormous.
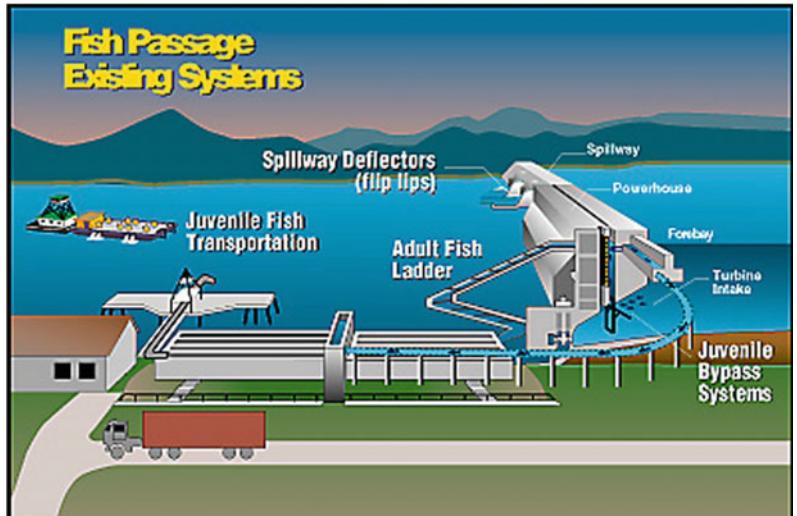
Over a dozen national and regional environmental organizations have joined together to bring back salmon and steelhead by modifying or partially dismantling five federal dams on the Columbia and Snake Rivers. Partial removal of the four dams on the lower Snake River in Washington State and lowering the reservoir behind John Day dam on the Columbia bordering

Fig. 1.8 A salmon swimming upstream (**a**) and measures taken to protect young juvenile salmon pass by hydropower dams on their way downstream (**b**) (US Fish and Wildlife Service and US Army Corps of Engineers, Pacific region)

Oregon and Washington (see Fig. 1.8) should help restore over 200 miles of vital river habitat. Running the rivers more like rivers may return salmon and steelhead to 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 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 help restore rivers, save salmon, and return balance to the Northwest's major rivers.

1.2.5 Wetland Preservation: A Groundswell of Support and Criticism

The balmy beach community of Tiger Point near Pensacola, Florida, bordering the Gulf of Mexico, is booming with development. New subdivisions, a Wal-Mart discount retail store and a recreation center dot the landscape.

Most—if not all—of this neighborhood was once a wetland that soaked up rain during downpours. Now, water runs off the parking lots and the roofs and into resident's living rooms. Some houses get flooded nearly every year.

A federal agency oversees wetland development. Critics say the agency is permitting in this area one of the highest rates of wetland loss in the nation. Obviously local developers wish they did not have to deal with the agency at all. The tension in Tiger Point reflects the debate throughout the US about whether the government is doing enough—or too much—to protect the nation's environment, and in this case, its wetlands.

Environmentalists and some homeowners value wetlands because they help reduce water pollution and floods, as well as nurture a diverse wildlife population. But many landowners and developers see the open wetlands as prime territory for building houses and businesses, rather than for breeding mosquitoes. They view existing federal wetland rules as onerous, illogical, and expensive.

While some areas such as Tiger Point have residents who want stricter laws to limit wetlands development, others—such as the suburbs around Seattle—have people who long for less strict rules.

Federal regulators had tried to quell the controversy with a solution known as wetlands mitigation. Anyone who destroys a wetland is required to build or expand another wetland somewhere else. Landowners and developers also see mitigation as a way out of the torturous arguments over wetlands. However, studies have shown many artificial marshes do not perform as well as those created by nature (NRC 2001). Many of the new, artificial wetlands are what scientists call the "ring around the pond" variety: open water surrounded by cattails. Furthermore, the federal agency issuing permits for wetland replacement do not have the resources to monitor them after they are approved. Developers know this.

1.2.6 Lake Source Cooling: Aid to Environment, or Threat to Lake?

It seems to be 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 (Fig. 1.9). No more chlorofluorocarbons, the refrigerants that can destroy protective ozone in the atmosphere and at a cost substantially smaller than for conventional air conditioners. The lake water is returned to the lake, with a few added calories.

Fig. 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)

However, a group of local opponents insists that Cornell University's $55 million lake-source-cooling plan that replaced its aging air conditioners is actually an environmental threat. They believe it could foster algal blooms. Pointing to 5 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.

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 that took place in getting the required local, state, and federal permits approved. This is in spite of the fact that the planning process, that took over 5 years, requested and involved the participation of all interested stakeholders (that would participate) from the very beginning. Even the local Sierra Club chapter 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 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're 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 a decade, and so far no adverse environmental effects have been noticed at any of the many monitoring sites.

1.2.7   Managing Water in the Florida Everglades

The Florida Everglades (Fig. 1.10) is the largest single wetland in the continental United States. In the mid-1800s it covered a little over nine million acres, but since that time the historical Everglades has been drained and half of the area 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 nonnative or otherwise uncommon plants and animals have out-competed native species. Populations of wading birds (including some endangered species) have declined by 85–90% 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 (Fig. 1.11) will not be able to provide adequate water supplies to agricultural and urban areas, or sufficient flood protection, 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 because of variations from the natural/historical hydrologic 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 Everglades' aquatic ecosystem to a more natural regime, an ambitious plan has been developed by the U.S. Army Corps of Engineers 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 nongovernmental organizations, as well as from the public at large.

The plan to restore the Everglades is ambitious and comprehensive, involving change of the current hydrologic regime in the remnant Everglades to one that resembles a more natural one, reestablishment of marshes and wetlands, implementation of agricultural best management practices, enhancements for wildlife and recreation, and provisions for water supply and flood control.

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 that are available. They also require the political will to make compromises and to put up with the lawsuits by anyone possibly disadvantaged by some restoration measure.

Who pays for all this? The taxpayers of Florida and the taxpayers of the U.S.

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

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

1.2.8 Restoration of Europe's Rivers and Seas

1.2.8.1 North and Baltic Seas

The North and Baltic Seas (shown in Fig. 1.12) are the most densely navigated seas in the world. Besides shipping, military, and recreational uses, an offshore oil 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.

Being so intensively used and surrounded by advanced industrialized countries, pollution problems are serious. The main pollution sources include various wastewater outfalls, dumping by ships (of dredged materials, sewage sludge, and chemical wastes) and operational discharges from offshore installations. 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 mammals and fish species and the sea grass community. 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 accumulate in the sediments and in the food web.

Fig. 1.12 Europe's major rivers and seas

Some animals, such as the gray seal and the sea eagle, are threatened with extinction.

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

1.2.8.2 The Rhine

The map of Fig. 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 2 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 2 months of the same year there was hardly enough water in the 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.
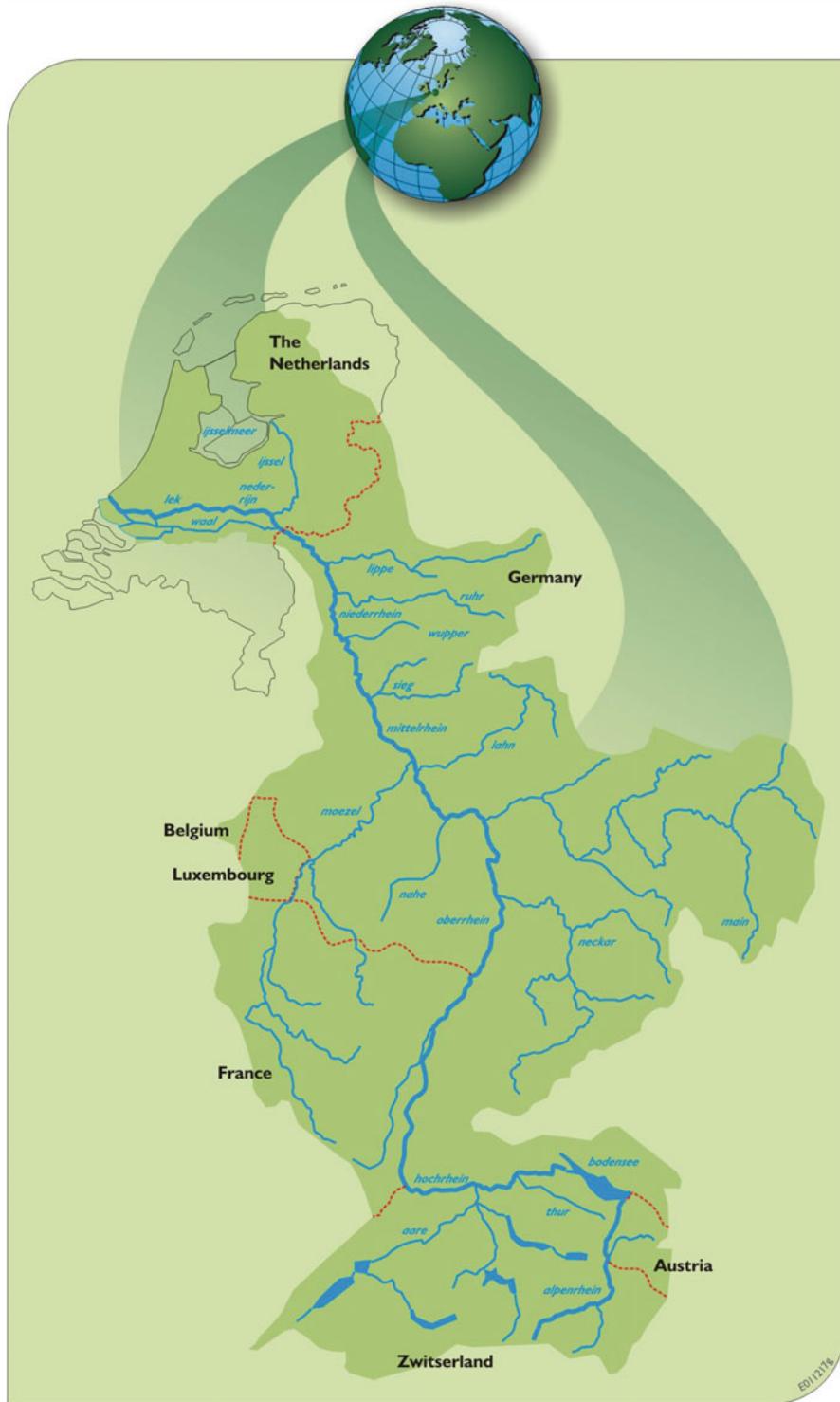
To address just how to restore the Rhine, it is useful to look at what has been happening to the river during the past 150 years. The Rhine was originally a natural watercourse. It is the only river connecting the Alps with the North Sea. To achieve greater economic benefits from the river, it was engineered for navigation, hydropower, water supply, and flood protection. Flood plains now "protected" from floods, provided increased land areas suitable for development. The main stream of the Rhine is now considerably shorter and 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. The basin is characterized by intensive industrial and agricultural activities. Some 20% of the world's chemical industry is located in the Rhine River basin. The River is reportedly the busiest shipping waterway in the world, containing long canals with regulated water levels. These canals connect the Rhine and its tributaries with the rivers of almost all the surrounding river basins including the Danube River. 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 growing from the recent toxic spill and floods as from a generally increasing interest by the inhabitants of the basin in environmental and ecosystem restoration and the preservation of natural beauty, has resulted in basin-wide efforts to rehabilitate the basin to a more "living" sustainable entity.

A Rhine Action Programme was 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 the revival of ecologically important areas along the Rhine. The plan, implemented in the 1990s, was given the name "Salmon 2000". 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 pose a challenge, as no one pretends that the

Fig. 1.13 The Rhine River Basin of Western Europe and its extension in The Netherlands

engineering works that provide navigation and hydropower benefits, but which also inhibit fish passage, are no longer needed or desired.

1.2.8.3 The Danube

The Danube River (shown in Fig. 1.14) is in the heartland of Central Europe. Its basin includes to a larger extent the territories of 15 countries. It additionally receives runoff from small catchments located in four other countries. About 90 million people live in the basin. This river encompasses perhaps more 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 River provides a vital resource for drainage, communications, transport, power generation, fishing,

recreation, and tourism. It is considered to be an ecosystem with irreplaceable environmental values.

More than 40 dams and large barrages plus over 500 smaller reservoirs have been constructed on the main Danube River and its tributaries. Flood control dikes confine most of the length of the main stem of the Danube River and the major tributaries. Over the last 50 years natural alluvial flood plain areas have declined from about 26,000 km^2 to about 6000 km^2.

There are also significant reaches with river training works and river diversion structures. These structures trap nutrients and sediment in the reservoirs. This 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

Fig. 1.14 The Danube River in Central Europe

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 nonpoint source agricultural, industrial, and municipal wastes. Because of the poor water quality (sometimes affecting human health) the riparian countries of the Danube river 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 far as possible adverse impacts and changes occurring or likely to be caused."

1.2.9   Flood Management on the Senegal River

As on many rivers in the tropical developing world, dam constructions on the Senegal (and conventional dam management strategies) can change not only the riverine environment but also the social interactions and economic productivity of farmers, fishers, and herders whose livelihoods depend on the annual flooding of valley bottomlands. Although much of the Senegal River flows through a low rainfall area, the naturally occurring annual flooding supported a rich and biologically diverse ecosystem. Living in a sustainable relationship with their environment, small-land holders farmed sandy uplands during the brief rainy season, and then cultivated the clay plains as floodwaters receded to the main channel of the river. Livestock also benefited from the succession of rain-fed pastures on the uplands and flood-recession pastures on the plains. Fish were abundant. As many as 30,000 tons were caught yearly. Since the early 1970s, small irrigated rice schemes added a fifth element to the production array: rain-fed farming, recession farming, herding, fishing, and irrigation.

Completion of the Diama salt intrusion barrage near the mouth of the river between Senegal and Mauritania and Manantali High Dam more than 1000 km upstream in Mali (Fig. 1.15), and the termination of the annual flood have had adverse effects on the environment. Rather than insulating the people from the ravages of drought, the dam release policy can accelerate desertification and intensify food insecurity. Furthermore, anticipation of donor investments in huge irrigation schemes has, in this particular case, lead to the expulsion of non-Arabic-speaking black Mauritanians from their floodplain lands.

Fig. 1.15 Senegal River and its Manantali Reservoir more than 1000 km upstream in Mali

This is a common impact of dam construction: increased hardships of generally politically powerless people in order that urban and industrial sectors may enjoy electricity at reduced costs.

Studies in the Senegal Valley by anthropologists, hydrologists, agronomists, and others suggest that it may be entirely economically feasible to create a controlled annual "artificial flood," assuring satisfaction of both urban, industrial, and rural demands for the river's water and supporting groundwater recharge, reforestation, and biodiversity.

Because of these studies, the government of Senegal ended its opposition to an artificial flood, and its development plans for the region are now predicated on its permanence. However, due to the common belief that releasing large quantities of water to create an artificial flood is incompatible with maximum hydropower production, the other members of the three-country consortium managing the dams—Mali and Mauritania—have resisted accepting this policy.

1.2.10   Nile Basin Countries Striving to Share Its Benefits

The Nile River (Fig. 1.16) is one of the major rivers of the world, serving millions and giving birth to entire civilizations. It is one of the world's longest rivers, traversing about 6695 km from the farthest source of its headwaters in Rwanda and Burundi through Lake Victoria, to its delta in Egypt on the Mediterranean Sea. Its basin includes 11 African countries (Burundi, DR Congo, Egypt, Eritrea, Ethiopia, Kenya, Rwanda, South Sudan, The Sudan, and Tanzania) and extends for more than 3 million square kilometers which represents about 10% of Africa's land mass area. The basin includes the Sudd wetland system in South Sudan.

Nile Basin countries are today home to more than 437 million people and of these, 54% (238 million) live within the basin and expect benefits from the management and use of the shared Nile Basin water resources.

Notwithstanding the basin's natural and environmental endowments and rich cultural history, its people face considerable challenges including persistent poverty with millions living on less than a dollar a day; extreme weather events associated with climate variability and change such as floods and droughts; low access to water and sanitation services; deteriorating water quality; and very low access rate to modern energy with most countries below 20% access level. The region also has a history of tensions and instability both between states and internal to states.

Cooperative management and development could bring a vast range of benefits including increased hydropower and food production; better access to water for domestic use; improved management of watersheds and reduced environmental degradation; reduced pollution and more control over damage from floods and droughts. Recognizing this the Nile Basin Initiative was created as a regional intergovernmental partnership that seeks to develop the River Nile in a cooperative manner, share substantial socioeconomic benefits, and promote regional peace and security. The partnership includes 10 Member States namely Burundi, DR Congo, Egypt, Ethiopia, Kenya, Rwanda, South Sudan, The Sudan, Tanzania, and Uganda. Eritrea participates as an observer. NBI was conceived as a transitional institution until a permanent institution can be created.

The partnership is guided by a Shared Vision: "To achieve sustainable socio-economic development through equitable utilization of, and benefit from, the common Nile Basin Water resources." The shared belief is that countries can achieve better outcomes for all the peoples of the Basin through cooperation rather than competition. It is supported by a "Shared Vision Planning Model" built by experts from all the basin countries. The model is designed to run different scenarios and assess the basin-wide impacts of different management policies and assumptions that any country may wish to perform.

Fig. 1.16 The Nile River Basin

1.2.11 Shrinking Glaciers at Top of the World

As shown in Fig. 1.17, Tibet lies north of India, Nepal, Bhutan, and Myanmar, west of China, and south of East Turkistan. The highest and largest plateau on Earth, it stretches some 1500 miles (2400 km) from east to west, and 900 miles (1448 km) north to south, an area equivalent in size to the United States region east of the Mississippi River. The Himalayas form much of its southern boundary, and Tibet's average altitude is so high—11,000 feet (3350 km) above sea level—that visitors often need weeks to acclimate.

The Tibetan Plateau serves as the headwaters for many of Asia's largest rivers, including the Yellow, Yangtze, Mekong, Brahmaputra, Salween, and Sutlej, among others. A substantial portion of the world's population lives in the watersheds of the rivers whose sources lie on the Tibetan Plateau.

Recent studies—including several by the Chinese Academy of Sciences—have documented a host of serious environmental challenges involving the quantity and quality of Tibet's freshwater reserves, most of them caused by industrial activities. Deforestation has led to large-scale erosion and siltation. Mining, manufacturing, and other human and industrial activities are producing record levels of air and water pollution in Tibet, as well as elsewhere in China
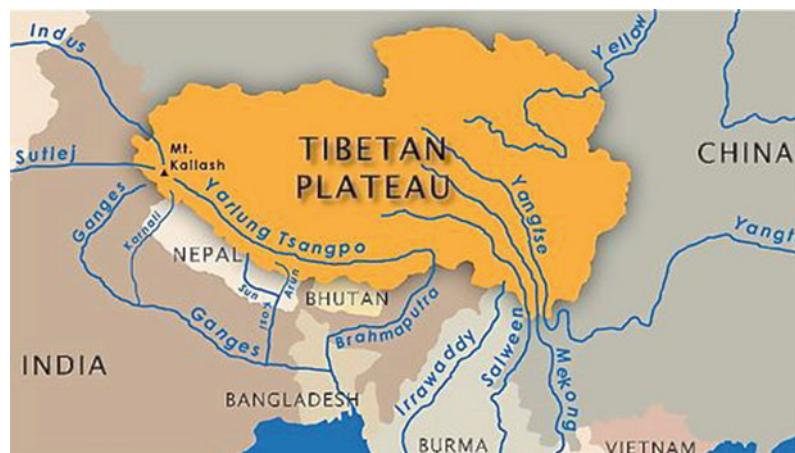
(Wong 2013). Together, these factors portend future water scarcity that could add to the region's political volatility.

Most important is that the region's glaciers are receding at one of the fastest rates anywhere in the world, and in some regions of Tibet by three 3 m per year (IPPC 2007). The quickening melting and evaporation is raising serious concerns in scientific and diplomatic communities, in and outside China, about Tibet's historic capacity to store more freshwater than anyplace on earth, except the North and South Poles. Tibet's water resources, they say, have become an increasingly crucial strategic political and cultural element that the Chinese are intent on managing and controlling.

1.2.12 China, a Thirsty Nation

Why does China care about the freshwater in Tibet? With more than a quarter of its land classified as desert, China is one of the planet's most arid regions. Beijing is besieged each spring by raging dust storms born in Inner Mongolia where hundreds of square miles of grasslands are turning to desert each year. In other parts of the nation, say diplomats and economic development specialists, Chinese rivers are either too polluted or too filled with silt to provide all of China's people with adequate supplies of freshwater.

Fig. 1.17 China, India, and Southeast Asia, highlighting the Tibetan Plateau

Chinese authorities have long had their eyes on Tibet's water resources. They have proposed building dams for hydropower and spending billions of dollars to build a system of canals to tap water from the Himalayan snowmelt and glaciers and transport it hundreds of miles north and east to the country's farm and industrial regions.

But how long that frozen reservoir will last is in doubt. In attempting to solve its own water crisis, China could potentially create widespread water shortages among its neighbors.

While the political issues involving Tibet are complex, there is no denying that water plays a role in China's interest in the region. The water of Tibet may prove to be one of its most important resources in the long run—for China, and for much of southern Asia. Figuring out how to sustainably manage that water will be a key to reducing political conflicts and tensions in the region.

1.2.13 Managing Sediment in China's Yellow River

The scarcity of water is not the only issue China has to address. So is sediment, especially in the Yellow River (Fig. 1.18). The Yellow River basin is the cradle of Chinese civilization, with agricultural societies appearing on the banks of the river more than 7000 years ago. The Yellow River originates in the Qinghai–Tibetan plateau and discharges into the Bohai Gulf in the Yellow sea. The basin is traditionally divided into the upper, middle, and lower reaches, which can be described as three down-sloping steps: the Tibetan Plateau, the Loess Plateau, and the alluvial plain. Key management issues are many, but the most visible one is sediment (Figs. 1.19 and 1.20).

The high sediment load of the Yellow River is a curse if the sediment deposits on the bed of the channel and reduces its capacity, thereby increasing the risk of flooding. Also, rapid deposition of sediment in reservoirs situated along the river is a problem as it reduces their effectiveness for flood control and water storage.

Another major management issue is the ecosystem health of the river. The relative scarcity of water creates a tension between allocating water for the benefit of river health, and for direct social and economic benefit. Irrigation uses 80% of the water consumed from the river, with the rest supplying industry, and drinking water for cities along the river and outside of the basin (Tianjin, Cangzhou and Qingdao). During the 1980s and 1990s the lower river dried up nearly every year, resulting in lost cereal production, suspension of some industries, and insufficient water supplies for more than 100,000 residents, who had to queue daily for drinking water. As well as costing around RmB40 billion in lost production, there was a serious decline in the ecological health of the river.

The diversity of habitat types and extensive areas of wetlands within the Ramsar-listed Yellow River Delta support at least 265 bird species. The birds, fish, and macroinvertebrates in the delta rely on healthy and diverse vegetation communities, which in turn depend upon on annual freshwater flooding and the associated high sediment loads. Degradation of the ecosystem of the Delta has been documented, especially from the late-1990s, due to increased human activities and a significant decrease in the flow of freshwater to the Delta wetlands. This has led to saltwater intrusion and increased soil salinity. Restoration activities involving the artificial delivery of freshwater to the wetlands began in 2002.

1.2.14 Damming the Mekong (S.E. Asia), the Amazon, and the Congo

The world's most biodiverse river basins—the Amazon, Congo, and Mekong—are attracting hydropower developers. While hydropower projects address energy needs and offer the potential of a higher standard of living, they also can impact the river's biodiversity, especially fisheries. The Amazon, Congo, and Mekong basins hold roughly one-third of the world's freshwater

Fig. 1.18   The Yellow River Basin in China

fish species, most of which are not found else-
where. Currently more than 450 additional dams
are planned for these three rivers (see Figs. 1.22
and 1.23) (Winemiller et al. 2016). Many of the
sites most appropriate for hydropower production
also are the habitats of many fish species. Given
recent escalation of hydropower development in
these basins, planning is needed to reduce bio-
diversity loss, as well as other adverse environ-
mental, social, and economic impacts while
meeting the energy needs of the basins.

   The Mekong River (Fig. 1.21) flows some
4200 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 due to regional conflicts, indeed
conflicts that have altered the history of the world.
Now that these conflicts are not resulting in mil-
itary battles (at this writing), 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 sub-
stantial, but so are the environmental, ecological,
and social risks (Orr et al. 2012).

   The economic value of hydroelectric power
currently generated from the Mekong brings in
welcome income however the environmental
impacts are harder to quantify. Today some
60 million people (12 million households) live in
the Lower Mekong Basin, and 80% rely directly

Fig. 1.19 Sediment flows in China's Yellow River. http://yellowriver-china.blogspot.com/2011/09/book-review-on-flood-discharge-and.html

on the river system for their food and livelihoods. Most of these households would be affected by alterations to fish availability since fish is their main source of dietary protein. The food security impacts on these people due to the existing and proposed dam building and operation in Cambodia, Laos, Thailand, and Vietnam remain relatively unexplored. Dam builders have often failed to recognize, or wish to ignore, the crucial role of inland fisheries in meeting food security needs.
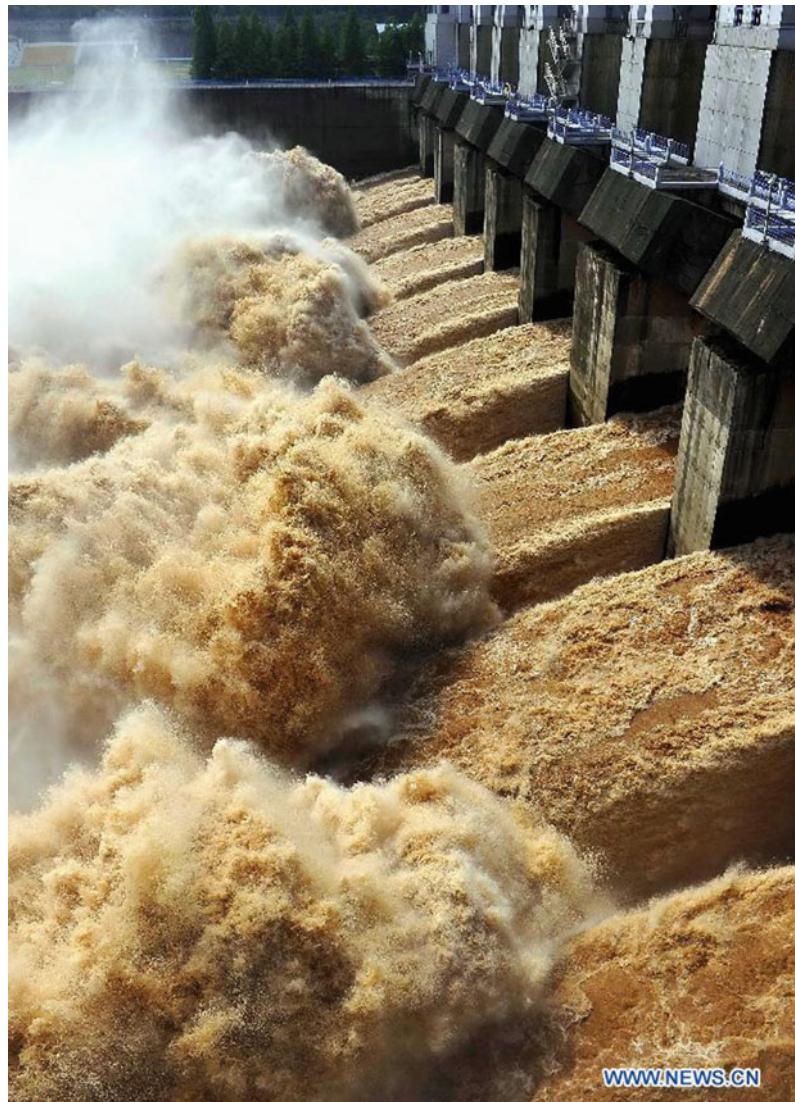
During some months of the year the lack of rainfall causes the Mekong to fall dramatically. Salt water may penetrate as much as 500 km inland. In other months the flow can be up to 30 times the low flows, causing the water in the river to back up into wetlands and flood some 12,000 km^2 of forests and 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 the inhabitants of the Mekong flood plain, 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.

What will happen to the social fabric and to the natural environment if the schemes to build big dams (see Fig. 1.22a) across the mainstream of the Mekong are implemented? Depending on their design, location, and operation, they could

Fig. 1.20 Dams can be designed and operated to remove some of the sediment that is trapped in the upstream reservoir

disrupt the current fertility cycles and the habitats and habits of the fish in the river resulting from the natural flow and sediment regimes. Increased erosion downstream from major reservoirs is also a threat. Add to these 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? And will there even be a Delta? Together with sea level rise and a blockage of Mekong's sediment to

the Delta, its survival as a geologic feature, and as a major source of food, is in doubt.

There have been suggestions that a proposed dam in Laos could cause deforestation in a wilderness area of some 3000 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 among

Fig. 1.21   The Lower Mekong River Basin including Tonle Sap Lake in Cambodia and the Mekong Delta in Vietnam

(a)

Mekong Mainstream Dams

Fig. 1.22 Lancang/Mekong River where reservoirs are being planned on the river itself (**a**) and on many of its tributaries (**b**). **a** http://khmerization.blogspot.com/2013/10/wwf-expresses-alarm-over-laos-decision.html, 6/10/13, and **b** reprinted from Wild and Loucks 2014, with permission. © 2014. American Geophysical Union

the types of issues that need to be considered now that increased development seems likely.

Similar issues face those who are planning similar hydropower dam developments in the other two most biodiverse river basins in the world—the Amazon and the Congo (Fig. 1.23). Clarifying the trade-offs between energy (economic), environmental, and social goals can inform governments and funding institutions as they make their dam siting, design, and operating decisions.

Hydropower accounts for more than two-thirds of Brazil's energy supply, and over 300 new Amazon dams have been proposed. Impacts of these dams would extend beyond direct effects on rivers to include relocation of

human populations and expanding deforestation associated with new roads. Scheduled for completion in 2016, Brazil's Belo Monte hydropower complex was designed with installed capacity of 11,233 MW, ranking it the world's third largest. But it could also set a record for biodiversity loss owing to selection of a site that is the sole habitat for many species. The Congo has far fewer dams than the Amazon or Mekong, yet most power generated within the basin is from hydropower. Inga Falls, a 14.5-km stretch of the lower Congo that drops 96 m to near sea level, has greater hydropower potential than anywhere else. The Inga I and II dams, constructed in the 1970s and 1980s, currently yield 40% of the 2132-MW installed capacity. Planned additional dams (Inga III and Grand Inga) would harness as much as 83% of the Congo's annual discharge, with most of the energy to be exported. Grand Inga would divert water and substantially reduce flow for at least 20 km downstream from the falls. Again, many trade-offs involved with dam building, and all calling for comprehensive systems planning and analyses to identify them.
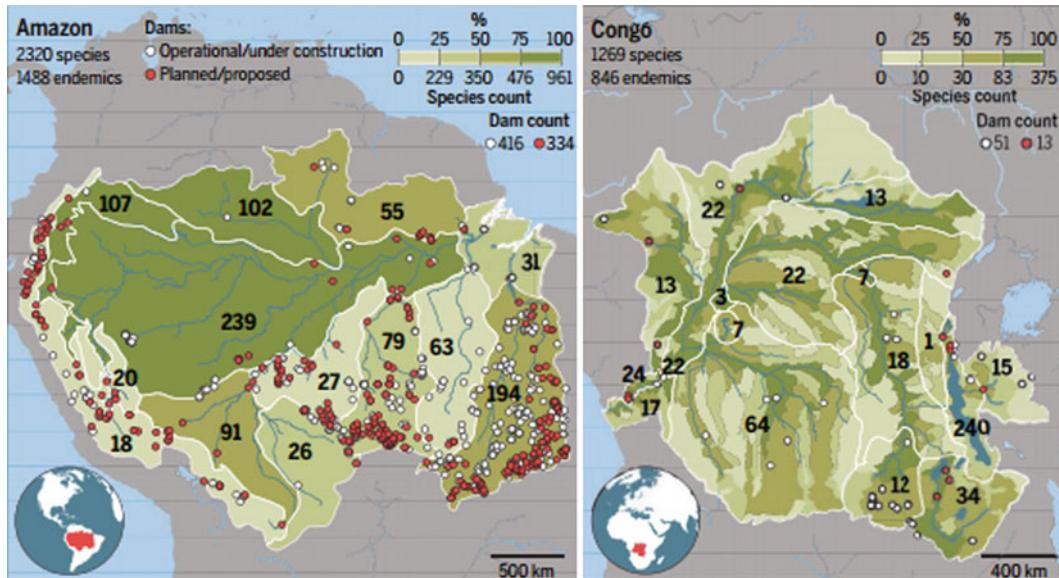
1.3 So, Why Plan, Why Manage?

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 problems to solve and/or opportunities to obtain increased benefits by changing the management and use of water and related land resources. These benefits can be measured in many different ways. The best way to do it is often not obvious. Whatever way 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 the best compromise 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 reasons include the identification and evaluation of alternative measures that may increase the available water supplies, hydropower, improve

(b)

Fig. 1.22 (continued)

Fig. 1.23 Fish diversity and dam locations in the Amazon and Congo basins. In addition to basin-wide biodiversity summaries (*upper left*), each basin can be divided into ecoregions (*white boundaries*). Approximate number of species (*black numbers*) and the total species richness (*shades of green*) found in ecoregions differ widely (Winemiller et al. 2016)

recreation and/or navigation, and enhance water quality 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 ranged 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; GTT 2014).

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. On the other hand, a reservoir used solely for downstream flood control is best left empty so it can store more of the flood flows when they occur. A single reservoir serving all three purposes introduces conflicts over how much water to store in it and discharge from it, i.e., 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 are reasons for planning.

1.3.1 Too Little Water

Issues involving inadequate supplies to meet demands can result from too little rain or snow. They can also result from patterns of land and water use. They can result from growing urbanization, the growing needs to meet instream flow requirements, and conflicts over private property and public rights regarding water allocations. Other issues can involve transbasin water transfers and markets, objectives of economic efficiency versus the desire to keep nonefficient 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 must cope with an actual water scarcity. The institutional authority
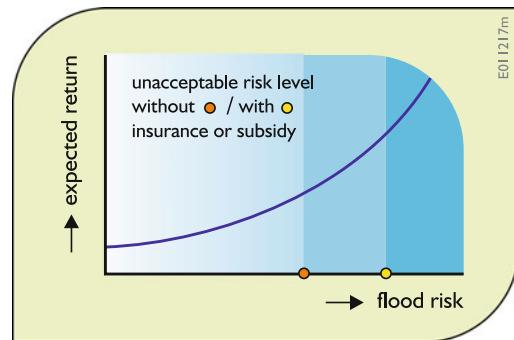
to implement drought measures when their designated "triggers"—such as storage volumes in reservoirs—have been met should be established before they are needed. Such management measures may include increased groundwater abstractions to supplement low-surface water flows and storage volumes. Conjunctive use of ground and surface waters can be sustainable as long as the groundwater aquifers are recharged during conditions of high flow and surface storage volumes. Many aquifers are subject to withdrawals exceeding recharge, and hence continued withdrawals from them cannot be sustained.

1.3.2 Too Much Water

Damage due to flooding is a direct result of floodplain development that is incompatible with 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 who incur losses in times of flooding. In many river basins of developed regions, annual expected flood damages are increasing over time, in spite of increased expenditures in flood damage reduction measures. This is in part due to increased economic development taking place on river flood plains, not only of increased frequencies and magnitudes of floods.

The increased economic value of developments on floodplains often justifies increased development and increased expenditures on flood damage reduction measures. Flood protection works decrease the risks of flood damage, creating an even larger 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 damages and costs of protection is a natural result of increasing values of flood plain development. Just what is the appropriate level of risk? It may depend, as Fig. 1.24 illustrates, on the level of flood insurance or subsidy provided when flooding occurs.

Flood damages will decrease only if there are restrictions placed on floodplain development.

Fig. 1.24 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

Analyses carried out during planning can help identify the appropriate level of development and flood damage protection works based on the beneficial as well as adverse economic, environmental, and ecological consequences of flood plain 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.

Industrial development and related port development may result in the demand for deeper and wider rivers to allow the operation of larger draft cargo vessels in the river. River channel improvement cannot be detached from functions such as water supply and flood control. Widening and deepening a river channel for shipping purposes may also decrease flood water levels.

1.3.3 Too Polluted

Wastewater discharges by industry and households can have considerable detrimental effects on water quality and hence on public and ecosystem health. Planning and management activities should pay attention to these possible negative consequences of industrial development and the intensive use and subsequent runoff 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,
- Nonpoint source pollution discharges including sediment from erosion, and
- 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, pharmaceutical products, and micropollutants and nanoparticles in our water supplies, water quality standards, plans and management policies should be adjusted accordingly. The occurrence of major fish kills and algae blooms also point to the need to manage water quality as well as quantity.

1.3.4 Too Expensive

Too many of the world's population do not have adequate water to meet all of their drinking and sanitation needs. Much of this is not due to the lack of technical options available to provide water to meet those needs. Rather those options are deemed to be too expensive. Doing so is judged to be beyond the ability of those living in poverty to pay and recover the costs of implementing, maintaining, and operating the needed infrastructure. Large national and international aid grants devoted to reducing water stress—demands for clean water exceeding usable supplies—in stressed communities have not been sustainable in the long run where recipients have been unable to pay for the upkeep of whatever water resource systems are developed and provided. If financial aid is to be provided, to be

effective it has to address all the root causes of such poverty, not only the need for clean water.

1.3.5 Ecosystem Too Degraded

Aquatic and riparian ecosystems may be subject to a number of threats. The most important ones 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. 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 nonindigenous 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 in prevention and early detection and eradication of aquatic nuisance species may

reduce the need to spend considerably more funds on management and control once such aquatic nuisance species are well established.

1.3.6 Other Planning and Management Issues

1.3.6.1 Navigation
Dredging river beds is a common practice to keep river channels open for larger draft cargo ships. The use of jetties as a way to increase the flow in the main channel and hence increase bottom scour is a way to reduce the amount of dredging that may be needed, but any modification of the width and depth of a river channel can impact its flood carrying capacity. It can also alter the periodic flooding of the floodplain that in turn can have ecological impacts.

1.3.6.2 River Bank Erosion
Bank erosion can be a serious problem where towns are located close to morphologically active (eroding) rivers. 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, buildings, and hydraulic structures.

1.3.6.3 Reservoir Related Issues
Degradation of the riverbeds 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. Reservoirs can be ecological barriers for migrating fish species such as salmon. The water quality in the reservoir may deteriorate and the inflowing sediment may settle and accumulate, reducing the active (useful) water storage capacity of the reservoir and causing more erosion downstream. 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 concentrations result in higher risks of scouring of downstream riverbeds and consequently a lowering of their elevations. Economic as well as social impacts include the risk of a dam break. Environmental impacts may result from sedimentation control measures (e.g., sediment flushing as shown in Fig. 1.19) and reduced oxygen content of the outflowing water.

1.4 System Planning Scales

1.4.1 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 impact 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 block vessels and fish from traveling up- or downstream through the dam site. To maximize the economic and social benefits obtained from the entire basin, and to insure that these benefits and accompanying costs are equitably distributed, planning and management on a basin scale is often undertaken.

While basin boundaries make sense from a hydrologic 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, social-economic, and administrative water resource system. To the extent that the applicable problems, stakeholders, and administrative boundaries extend outside the

river basin, then 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.

1.4.2   Temporal Scales for Planning and Management

Planning is a continuing iterative process. Water resources plans need to be periodically updated and adapt to new information, new objectives, and updated forecasts of future demands, costs, and benefits. Current decisions should not preclude future generations from options they may want to consider, but otherwise current decisions should be responsive to current needs and opportunities, and have the ability to be adaptable in the future to possible changes in those needs and opportunities.

The number and duration of within-year time periods explicitly considered in the planning process will depend in part on the need to consider the variability of the supplies of and demands for water resources and on the purposes to be served by the water resources. 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 typically take place on different spatial and temporal scales.

1.5   Planning and Management 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 the grassroots approach.

Both approaches, working together, can lead to an integrated plan and management policy.

1.5.1   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 nonstructural management options are identified and evaluated. Based on these evaluations, the preferred plan is recommended.

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

1.5.2   Bottom-Up Planning and Management

Within the past several decades water resources planning and management processes have increasingly involved the active participation of interested stakeholders—those potentially affected by the decision being considered. Plans are being created from the bottom-up rather than top-down through a process of consensus building. Concerned citizens, nongovernmental organizations, as well as professionals in

governmental agencies are increasingly working together toward the creation of adaptive comprehensive water management programs, policies, and plans.

Experiences trying to implement plans developed primarily by professionals without significant citizen involvement have shown that even if such plans are technically sound they have little chance of success if they do not take into consideration the concerns and objectives of affected stakeholders. To gain their support, concerned stakeholders must be included in the decision-making process as early as possible. They must become part of the decision-making process, not merely spectators, or even advisors, to it. This will help gain their cooperation and commitment to the plans eventually adopted. Participating stakeholders will consider the resulting plans as their plans as much as someone else's. They will have a sense of ownership, and as such will strive to make them work. Such adopted plans, if they are to be successfully implemented, must fit within existing legislative, permitting, enforcement, and monitoring programs. 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 and participate in the water resources planning and management process. It will involve building a consensus on goals and objectives and on how to achieve them. 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 does not get any easier.
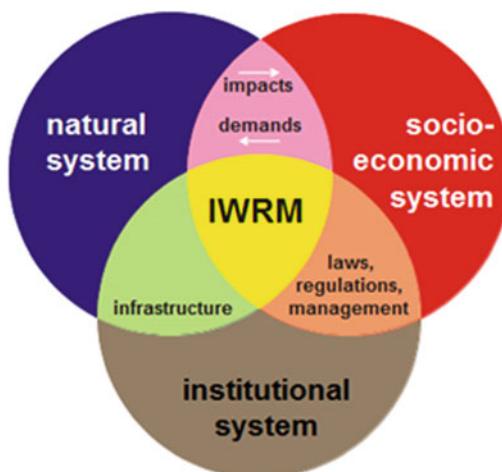
Bottom-up planning must strive to achieve a common or "shared" vision among all stakeholders. It must either comply with all applicable laws and regulations, or propose changes to them. It should strive to identify and evaluate multiple alternatives and performance criteria—including sustainability criteria, and yet keep the process from producing a wish list of everything each stakeholder wants. In other words, it must identify trade-offs among conflicting goals or measures of performance, and prioritizing appropriate strategies. It must value and compare, somehow, the intangible and nonmonetary impacts of environmental and ecosystem protection and restoration with other activities whose benefits and costs can be expressed in monetary units. In doing all this, planners should use modern information technology, as available, 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 emphasize the need to make water resources planning and management as efficient and effective as possible and remain participatory. 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 nonmonetary 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 and mobile Internet networks, decision support systems, etc.) can be beneficially used both for analyses as well as communication.

1.5.3   Integrated Water Resources Management

The concept of integrated water resources management (IWRM) has been developing over the past several decades. IWRM is the response to the growing pressure on our water resources systems caused by growing populations and socioeconomic developments. Water shortages and deteriorating water quality have forced many countries in the world to reconsider their development policies 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 toward a demand-oriented, multisectoral approach, often labeled integrated water resources management.

The concept of IWRM moves away from top-down "water master planning" that usually focuses on water availability and development, and toward "comprehensive water policy planning" that addresses the interaction between different subsectors (Fig. 1.25), seeks to establish priorities, considers institutional requirements, and deals with the building of management capacity.

Fig. 1.25 Interactions among the natural, administrative, and socioeconomic water resource subsectors and between them and their environment

Box 1.1 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)

IWRM (Box 1.1) considers the use of the resources in relation to social and economic activities and functions. These determine the need for laws and regulations pertaining to the sustainable and beneficial use of the water resources. Infrastructure together with regulatory measures allows more effective use of the resource including meeting ecosystem needs.

1.5.4   Water Security and the Sustainable Development Goals (SDGs)

While IWRM focuses on the *process* to improve water management (the how), the term "water security" focuses on the *output* (the what). The World Economic Forum has identified Water Security as one of the biggest global economic development issues. Water Security is defined by UN-Water (2013) as

> the capacity of a population to safeguard sustainable access to adequate quantities of acceptable quality water for sustaining livelihoods, human well-being, and socio-economic development, for ensuring protection against water-borne pollution and water-related disasters, and for preserving ecosystems in a climate of peace and political stability.

Attempts are being made to identify the many dimensions of water security and to quantify them (van Beek and Arriens 2014; ADB 2016). In 2015 the UN adopted the Sustainable Development Goals 2015–2030 that specify specific

targets for various goals such as the provision of water for drinking and sanitation, water productivity in agriculture, industry and energy, environment, and reduction of floods and droughts. It is expected that many countries will expect their water managers to use the SDGs as objectives in water resources planning. This means that our planning and management proposals need to be able to quantify the impacts of possible plans and policies in terms of the SDG targets.

1.5.5 Planning and Management Aspects

1.5.5.1 Technical

Technical aspects of planning include hydrologic assessments. Hydrologic assessments identify and characterize the properties of, and interactions among, the resources in the basin or region. This includes 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.) in the runoff resulting from any given pattern of rainfall over the land area. These predictions will help planners estimate the quantities and qualities of flows 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 impact the economic development of the region, which is what, 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. By storage of floodwater they may also reduce flood damage below the reservoir.
- 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 confining flows within predetermined channels.

Not only must the planning process identify and evaluate alternative management strategies involving structural and nonstructural 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 all impact 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 or changes in flow and sediment patterns due to upstream reservoir operation.
- decline in number and extent of wetlands and the adverse impacts to wetlands of proposed land and water development projects.
- conflicts between the needs of people for water supply, recreational, 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 balance these often-conflicting interests.

1.5.5.2   Financial and Economic

The overriding financial component of any planning process is to make sure that the recommended plans and projects will be 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. 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 how this is to be achieved should be taken into account. Yet beneficiaries should be expected to pay at least something for the added benefits they get. Planning must identify equitable cost and risk-sharing policies and improved approaches to risk/cost management.

Financial viability is often 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 nonmonetary objectives. In many developing countries a distinction is made between the recovery of investment costs and the recovery of O&M costs. 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 over time.

Many past failures in water resources management are attributable to the fact that water—its quantity, reliability, quality, pressure, location—has been and still is viewed as a free good. Prices paid 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 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 an inefficient use of the resource.

1.5.5.3   Institutional and Governance

The first condition for the successful implementation of plans and policies is to have an enabling environment. There must exist national, provincial, and local policies, legislation and institutions that make it possible for the desired decisions to be taken and implemented. 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 investments to develop, treat, store, distribute,

and use, and then to collect, treat, and dispose or reuse. Examples are multipurpose reservoirs and the construction of dykes along coasts and rivers. The required investments are large and typically can only be made by governments or state-owned companies.

- Water is a medium that can easily transfer external effects. The use of water by one activity often has negative effects on other water using activities (externalities). The obvious example is the discharge of wastewater into a river may save the discharger money but it may have negative effects on downstream users requiring cleaner water.

Only the government can address many of these issues and hence "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 and governance in both developed and developing regions and 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.

1.5.5.4 Models for Impact Prediction and Evaluation

Planning processes have undergone a significant transformation over the past five decades, mainly due to the continuing development of improved computational technology. 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 using 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 following 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 models 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 to 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 modeling 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 modeling programs or "shells" into which interested stakeholders can "draw in" their system, enter their data and operating rules at the level of detail desired, simulate it, 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.

1.5.5.5 Models for Shared Vision or Consensus Building

Participatory planning involves conflict management. Each stakeholder or interest group has its objectives, interests, and agendas. Some of these may be in conflict. 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 the right decisions 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 modeling process.

Involving stakeholders in collaborative 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 do. If they are involved in model building, they will know the assumptions built into their model.

Being involved in a modeling exercise is a way to understand better the impacts of various assumptions one must make when developing and running models. 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 involvement of stakeholders in the process of model development will create discussions that will lead toward a better understanding of everyone's interests and concerns. Though 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 system (as represented by their model, of course) works.

1.5.5.6 Models for Adaptive Management

Recent emphasis has shifted from structural engineering solutions to more nonstructural 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 based on the results of an optimization or simulation model of a particular water resource system 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 modeling software used to estimate various impacts, collect and analyze new data, and maintain the expertise, all of which are necessary for continuous planning (rolling plans)?

1.6 Planning and Management Characteristics

1.6.1 Integrated Policies and Development Plans

Clearly, a portion of any water resources planning and management study report should contain a discussion of the particular site-specific water resource 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 nonstructural 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 their

water resources, we do not want to be regretting what we have done in the past that may preclude this adaptation.

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 hydrologic, ecologic, 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 as mitigating for 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 centers 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 trade-offs associated with different areas of different 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 with 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 ecologic benefits of having light residential zones between waterways and commercial, urban, or agriculture 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 agricultural land. 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 forests 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, i.e., in multiple river basins, we just might begin to understand and predict better the interactions among economy, environment ecology, and people 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 on into the future.

1.6.2   Sustainability

Sustainable water resource systems are those designed and managed to best serve people living in the future as well as those of us living today. The actions that we as a society take now to satisfy our own needs and desires should not only depend 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 water resource systems that we judge best serves those of us living today as well as our children and their children living in the future.

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 do not 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 trade-offs 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 trade-offs, if any, between what present and future generations would like should be considered. Once identified, or at least estimated, just what decisions to make should be debated and decided in the political arena. There is no scientific theory to help us identify which trade-offs, 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 impact 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, to aging of infrastructure, to shifts in demands or desires of a changing society, and even due to 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 resource systems are those designed and operated in ways that make them more adaptive, robust, and resilient to an uncertain and changing future. Sustainable water resource systems must be capable of effectively functioning 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 and new objectives and new 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 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 and 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 socioeconomic and natural environments.

Changing the social and institutional components of water resource systems are often the most challenging because they involve 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 that allow future generations to alter such systems. 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 resource 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 planning and management project.

1.7 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. Planning 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 how the resources are to be allocated or controlled, especially in times of stress, and
- 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: (1) a growing concern for the sustainability of natural ecosystems and (2) an increased recognition for the need of the bottom-up "grassroots" 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 have done in the past, 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 hydrologic regimes. Other users prefer less variability. Much of our engineering infrastructure is operated so as to reduce hydrologic variability. Today water resource systems are increasing, required to provide rather than reduce hydrologic (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 requiring habitats along the full length of rivers to complete their life cycles must have access to those habitats. 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 program administration, and broad participation of stakeholders. Yet obtaining each of these three conditions is a difficult 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 are 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, nongovernmental 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 its own 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 just 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 gets 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,
- 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, or 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 to just criticize it from the outside? Who is in a position at the local level to provide that leadership and needed financial support? In some regions, nongovernmental institutions have been instrumental in initiating and coordinating this process at local grassroot 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, and
- 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 first step toward becoming an accomplished planner or manager.

References

ADB (Asia Development Bank). (2016). *Asian Water Development Outlook, Strengthening Water Security in Asia and the Pacific*. Mandaluyong City, Philippines.

ASCE (American Society of Civil Engineers). (1998). *Sustainability criteria for water resource systems*. Reston, VA: ASCE.

Gulbenkian Think Tank (GTT). (2014). *Water and the future of humanity*. Revaluating Water Security, Calouste Gulbenkian Foundation, New York: Springer. ISBN: 978-3-319-01456-2. doi:10.1007/978-3-319-01457-9_1

GWP (Global Water Partnership). (2000). *Integrated water resources management*. TAC Background Papers No. 4. Stockholm, Sweden: GWP.

IPPC. (2007). *Report of the intergovernmental panel on climate change*, May.

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.

Orr, S., Pittock, J., Chapagain, A., & Dumaresq, D. (2012). Dams on the Mekong River: Lost fish protein and the implications for land and water resources, *Global Environmental Change, 22*, 925–932.

UN-Water. (2013). *UN-water analytical brief on water security and the global water agenda*.

van Beek, E., & Arriens, W. L. (2014). *Water security, putting the concept into practise*. TEC Background Paper No. 20. Stockholm: GWP.

Winemiller, K. O., McIntyre, P. B. , Castello, L., Fluet-Chouinard, E., Giarrizzo, T., Nam, S., et al. (2016). Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong Basin-scale planning is needed to minimize impacts in mega-diverse rivers. In Insights perspectives (Vol. 351 Issue 6269 p. 129). AAAS Science. science-mag.org. January 8, 2016.

Additional References (Further Reading)

Abdullah, A., Masih, I.,Van Der Zaag, P., Karim, U. F.A., Popescu, I., Suhail, Q. A. (2015). Shatt al Arab River system under escalating pressure: A preliminary exploration of the issues and options for mitigation. *International Journal of River Basin Management, 13*(2), 215–227.

Abu-Zeid, M. A., & Biswas, A. K. (Eds.). (1996). *River basin planning and management*. Calcutta: Oxford University Press.

American Society of Civil Engineers (ASCE). (1998). *Sustainability criteria for water resource systems*. Reston, VA: ASCE Press.

American Society of Civil Engineers (ASCE). (2011). *Toward a sustainable water future: Visions for 2050*. Reston, VA, USA: ASCE Press.

Barrow, C. J. (1998). River basin development planning and management: A critical review. *World-Development, Oxford, United Kingdom, 26*(1), 171–186.

Bateman, B., & Rancier, R. (Eds.). (2012). *Case studies in integrated water resources management: From local stewardship to national vision*. American Water Resources Association Policy Committee, 4 West Federal Street, P.O. Box 1626, Middleburg, Virginia 20118-1626, 60 pp.

Bates, B. C., Kundzewicz, Z. W., Wu, S., & Palutikof, J. P. (2008). *Climate change and water*. IPCC Technical Paper VI. Geneva, Switzerland: Intergovernmental Panel on Climate Change.

Biswas, A. K. (Ed.). (1997). *Water resources: Environmental planning, management, and development* (738 pp). New York, NY: McGraw-Hill.

Cai, X., Vogel, R., & Ranjithan, R. (Eds.). (2013). Role of systems analysis in watershed management. *Journal of Water Resources Planning and Management, 139*(5), 461–592.

Chartres, C., & Varma, S. (2010). *Out of water: From abundance to scarcity and how to solve the world's water problems*. Upper Saddle River, NJ, USA: Pearson Education Inc.

Cooper, A. B., & Bottcher, A. B. (1993). Basin-scale modeling as a tool for water-resource planning. *Journal of Water Resources Planning and Management (ASCE), 119*(3), 306–323. May/June 6 fig, 4 tab, 41 ref.

Priscoli, J. D. (2004). What is public participation in water resources management and why is it important? *Water International, 29*(2).

Diamantini, C., & Zanon, B. (1996). River basin planning in Italy: Resource and risk management. *European-Environment (United Kingdom), 6*(4), 119–125.

Eckstein, O. (1958). *Water resource development: The economics of project evaluation*. Cambridge, MA: Harvard University Press.

Frederick, K. D., Major, D. C., & Stakhiv, E. Z. (1997). Water resources planning principles and evaluation criteria for climate change: Summary and conclusions. *Climate Change, 37*(1), 291–313.

Gershon, M., & Duckstein, L. (1983). Multiobjective approaches to river basin planning. *Journal-Water-Resources-Planning-Management-Division, American Society of Civil Engineers USA), 109*(1), 13–28.

Gleick, P. H. (2002). Soft water paths. *Nature, 418*, 373.

Gleick, P. H. (2003). Water use. *Annual Review of Environment and Resources, 28*, 275–314.

Gleick, P. H., & Palaniappan, M. (2010). Peak water: Conceptual and practical limits to freshwater withdrawal and use. *Proceedings of the National Academy of Sciences of the United States of America, 107*(25), 11155–11162.

Global Water Partnership (GWP). (2000). *Water as a social and economic good: How to put the principle into practice*. TAC Background Papers No. 2. Stockholm, Sweden: GWP.

Global Water Partnership (GWP). (2000). *Effective water governance*. TAC Background Papers No. 7. Stockholm, Sweden: GWP.

Global Water Partnership (GWP). (2012). *The handbook for integrated water resources management in transboundary basins of rivers, lakes and aquifers* (120 pp). International Office for Water (Permanent Technical Secretariat of the International Network of Basin Organizations). ISBN: 978-91-85321-85-8.

Goulter, I. C. (1985). Equity issues in the implementation of river basin planning. *Strategies for river basin management: Environmental integration of land and water in a river basin* (pp. 287–292, 15 refs.). Dordrecht, Holland: D. Reidel Publishing Co.

Gulbenkian Think Tank. (2014). *Water and the future of humanity. Revisiting water security* (241 pp). New York: Springer. doi:10.1007/978-3-319-01457-9

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change, 23*, 485–498.

Habersack, H., Haspel, D., & Campbell, I. (Eds.). (2013). Integrated management of large river systems. *International Journal of River Basin Management, 11*(2), 137–236.

Hoekstra, A. Y., & Chapagain, A. K. (2008). *Globalization of water. Sharing the planet's freshwater resources*. Oxford, UK: Blackwell Publishing.

Hoekstra, A. Y., & Mekonnem, M. M. (2012). The water footprint of humanity. *Proceedings of the National Academy of Sciences, Washington DC, 109*(9), 3232–3237.

Howe, C. W. (1996). Water resources planning in a federation of states: Equity versus efficiency. *Natural Resources Journal, 36*(1), 29–36.

Huaicheng, Guo, & Beanlands, G. (1994). A comparative study on Canadian and Chinese river basin planning. *Journal of Environmental Science China, 6*(2), 224–233.

IPCC (Intergovernmental Panel on Climate Change). (2001). *Climate change 2001: Synthesis report, summary for policymakers*. Third Assessment Report, IPCC.

IPCC (Intergovernmental Panel on Climate Change). (2007). *Climate change 2007: The physical science basis.* Fourth Assessment Report, IPCC.

IWMI (International Water Management Institute). (2007). *Water for food, water for life. A comprehensive assessment of water management in agriculture*. London, UK: Earthscan, and Colombo, Sri Lanka: IWMI.

Karamouz, M., Szidarovszky, F., & Banafsheh, Z. (2003). *Water resources systems analysis*. Boca Raton, FL: Lewis.

Kulshreshtha, S. (1998). A global outlook for water resources to the year 2025. *Water Resources Management, 12*(3), 167–184.

Krutilla, J. V., & Eckstein, O. (1958). *Multiple purpose river development*. Baltimore, MD: Johns Hopkins Press.

Lee, D. J., & Dinar, A. (1996). Integrated models of river basin planning, development, and management. *Water International, 21*(4), 213–222. Also see: (1995). *Review of integrated approaches to river basin planning, development and management*. World Bank, Agriculture and Natural Resources Department, Washington, DC.

Lins, H. F., Wolock, D. M., & McCabe, G. J. (1997). Scale and modeling issues in water resources planning. *Climate Change, 37*(1), 63–88.

Loucks, D. P., Stedinger, J.R., & Haith, D.A. (1981). *Water resources systems planning and analysis* (559 pp). Englewood Cliffs, NJ: Prentice-Hall.

Loucks, D. P., & da Costa, J. R. (Eds.). (1991). *Decision support systems: Water resources planning and research*. Berlin: Springer.

Loucks, D. P. (Ed.). (1998). *Restoration of degraded rivers: Challenges, issues and experiences* (484 pp). Dordrecht, NL: Kluwer Academic Publishers.

Maidment, D. R. (Ed.). (1993). *Handbook of hydrology*. New York, NY: McGraw-Hill.

Maass, A., et al. (1962). *Design of water resource systems*. Cambridge, MA: Harvard University Press.

Major, D. C., & Lenton, R. L. (Eds.). (1979). *Applied water resource systems planning* (248 pp). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Mays, L. W. (Ed.). (1996). *Water Resources Handbook*. NY, NY: McGraw-Hill.

McMillan, T. (1990). Water resource planning in Canada. *Journal of Soil and Water Conservation, 45*(6),614–616.

Mimikou, M. A., Baltas, E. A., & Tsihrintzis, V. A. (2016). *Hydrology and water resource systems analysis* (456 pp). CRC Press.

Mitchell, B. (1983). Comprehensive river basin planning in Canada: Problems and opportunities. *Water International, 8*(4), 146–153.

MRC (Mekong River Commission) Secretariat. (2009). *Inception report: MRC SEA for hydropower on the Mekong mainstream*. Phnom Penh, Cambodia: MRC and International Center for Environmental Management.

National Research Council (NRC). (2001). *Compensating for Wetland Losses Under the Clean Water Act* (348 pp). Committee on Mitigating Wetland Losses, Board on Environmental Studies and Toxicology, Water Science and Technology Board.

Nijhuis, M. (2015). Harnessing the Mekong, or killing it? *National Geographic, Journal of National Geographic Society, 227*(5), 102–129.

O'Riordan, J. (1981). New strategies for water resource planning in British Columbia. *Canadian Water Resources Journal, 6*(4), 13–43. 6 Figs, 5 Tabs.

Razavian, D.,Bleed, A. S., Supalla, R. J., & Gollehon, N. R. (1990). Multistage screening process for river basin planning. *Journal of Water Resources Planning and Management (ASCE), 116*(3), 323–334. May/June 3 fig, 1 tab, 19 refs, 3 append.

Reitsma, R. F., & Carron, J. C. (1997). Object-oriented simulation and evaluation of river basin operations. *Journal of Geographic Information and Decision Analysis, 1*(1), 9–24.

Reynolds, P. J. (1985). *Ecosystem approaches to river basin planning strategies for river basin management: Environmental integration of land and water in a river basin* (pp. 41–48, 1 fig, 18 refs). Dordrecht, Holland: D. Reidel Publishing Co.

Sadoff, C. W., Hall, J. W., Grey, D., Aerts, J. C. J. H., Ait-Kadi, M., Brown, C., et al. (2015). *Securing water, sustaining growth: Report of the GWP/OECD task force on water security and sustainable growth* (180 pp). UK: University of Oxford.

Saha, S. K., & Barrow, C. J. (Eds.). (1981). *River basin planning: Theory and practice* (357 pp). Chichester, UK: Wiley Interscience.

Savenije, H. H. G., & van der Zaag, P. (Eds.). (1998). *The management of shared river basins* (164 pp). The Hague, NL: Ministry of Foreign Affairs, Neda.

Schramm, G. (1980). Integrated river basin planning in a holistic universe. *Natural Resources Journal, 20*(4), 787–806. 2 Fig, 1 tab, 48 refs.

Smith, S. C., & Castle, E. N. (Eds.). (1964). *Economics and public policy in water resources development*. Ames, Iowa: Iowa University Press.

Somlyody, L. (1997). Use of optimization models in river basin water quality planning. In M. B. Beck & P. Lessard (Eds.), *WATERMATEX '97. Systems analysis and computing in water quality management. Towards a new agenda* (391 pp).

Stout, G.E. (1998). Sustainable development requires the full cooperation of water users. *Water International, 23*(1), 3–7.

Thanh, N. C., & Biswas, A. K. (Eds.). (1990). *Environmentally-sound water management* (276 pp). Delhi: Oxford University Press.

Thiessen, E. M., Loucks, D. P., & Stedinger, J. R. (1998). Computer-assisted negotiations of water resources conflicts. *Group Decision and Negotiation, 7*(2).

Thissen, W. A. H., & Walker, W. E. (Eds.). (2013). *Public policy analysis: New developments* (286 pp). New York, Heidelberg, Dordrecht, London: Springer. ISBN: 978-1-4614-4601-9; ISBN: 978-1-4614-4602-6 (eBook). doi:10.1007/978-1-4614-4602-6

Tolley, G. S., & Riggs, F. E. (Eds.). (1961). *Economics of watershed planning*. Ames, Iowa: Iowa State University Press.

UN WWAP (United Nations World Water Assessment Programme). (2009). *The United Nations World Water Development Report 3: Water in a changing world*. Paris, France: UNESCO and London, UK: Earthscan.

UN WWAP (United Nations World Water Assessment Programme). (2011). *World water scenarios to 2050, exploring alternative futures of the world's water and its use to 2050*. Paris, France: UNESCO.

UN WWAP (United Nations World Water Assessment Programme). (2012). *The United Nations World Water Development Report 4: Managing water under risk and uncertainty*. Paris, France: UNESCO.

UNDP (United Nations Development Programme). (2011). Human Development Report. In *Sustainability and equity: A better future for all*. New York, NY, USA: UNDP.

UNEP (United Nations Environment Programme). (2002). *Global environmental outlook 3: Past, present and future perspectives*. London, UK: Earthscan and Nairobi, Kenya: UNEP.

UNEP (United Nations Environment Programme). (2006). *Marine and coastal ecosystems and human well-being: A synthesis report based on the findings of the millennium ecosystem assessment*. Nairobi, Kenya: UNEP.

UNEP (United Nations Environment Programme). (2009). *Towards sustainable production and sustainable use of resources: Assessing biofuels*. France: UNEP, produced by the International Panel for Sustainable Resource Management, Division of Technology Industry and Economics, France.

Viessman, W. (1996). Integrated water management. *Water Resources Update,* (106), 2–12.

Viessman, W. (1998). Water policies for the future. *Water Resources Update,* (111), 4–7, 104–110.

Vörösmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global water resources: Vulnerability from climate change and population growth. *Science, 289*(5477), 284–288.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., et al. (2010). Global threats to human water security and river biodiversity. *Nature, 467*, 555–561.

Walker, W. E., Haasnoot, M., & Kwakkel, J. H. (2013). Adapt or perish: A review of planning approaches for adaptation under deep uncertainty. *Sustainability, 5*, 955–979.

Watkins, D. W. Jr. (Ed.). (2013). *Water resources systems analysis through case studies: Data and models for decision making* (168 pp). Reston, VA: ASCE Press.

WCD. (2000). *Dams and developments—A new framework for decision-making*: *The report of the World Commission on Dams*. UK: Earthscan.

White, G. F. et al. (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.

Wong, E. (2013). Pollution rising, Chinese fear for soil and food. *NY Times*. December 31, 2013.

Wood, A. W., Lettenmaier, D. P., & Palmer, R. N. (1997). Assessing climate change implications for water resources planning. *Climate Change, 37*(1), 203–228.

Wright, W. C., Cohen, R., & Heath, J. H. (1982). Decentralizing water resource planning and management. *Journal of the American Water Works Association, 74*(7), 6, 334–345. July, 5 Fig, 2 Tab, 37 Ref.

WWAP, (United Nations World Water Assessment Programme). (2012). World water development report 4, 2012. *Managing water under uncertainty and risk* (867 pp). Paris: UNESCO.

WWAP (United Nations World Water Assessment Programme). (2015). World water development report 2015. *Water for a sustainable world* (224 pp; Case studies, 61 pp). Paris: UNESCO.

WWAP (United Nations World Water Assessment Programme). (2014). *The United Nations world water development report 2014: Water and energy* (133 pp). Paris: UNESCO.

Xie, J., Liebenthal, A., Warford, J. J., et al. (2009). *Addressing China's water scarcity: Recommendations for selected water resource management issues*. Washington, DC, USA: World Bank.

Zhang, R., Zhou, J., Zhang, H., Liao, X., & Wang, X. (2014). Optimal operation of large-scale cascaded hydropower systems in the Upper Reaches of the Yangtze River, China. *Journal of Water Resources Planning and Management, 140*(4), 480–495.

Exercises

1.1 How would you define "Integrated Water Resources Management" and what distinguishes it from "Sustainable Water Resources Management"?

1.2 Can you identify some common water management issues that are found in many parts of the world?

1.3 Comment on the common practice of governments giving aid to those in drought or flood areas without any incentives to alter land use management practices in anticipation of the next drought or flood.

1.4 What tools and information are available for developing integrated water resources plans and management policies?

1.5 What structural and nonstructural measures can be taken to address water resources issues?

1.6 Find the following statistics:
- Percent of all freshwater resources worldwide available for drinking;

- Number of people who die each year from diseases associated with unsafe drinking water;
- Percent of total freshwater resources in polar regions;
- Per capita annual withdrawal of cubic meters of freshwater in various countries;
- Average world per capita annual withdrawal of cubic meters of freshwater;
- Tons of pollutants entering lakes and rivers daily in various regions;
- Average number of gallons of water consumed by humans in a lifetime;
- Average number of kilometers per day a woman in a developing country must walk to fetch fresh water.

1.7 Identify and briefly describe the six greatest rivers in the world.

1.8. Identify some of the major water resource management issues in the region where you live. What management alternatives might effectively reduce some of the problems or provide additional economic, environmental, or social benefits.

1.9. Describe some water resource systems consisting of various interdependent natural, physical, and social components. What are the inputs to the systems and what are their outputs? How did you decide what to include in the system and what not to include?

1.10. Sustainability is a concept applied to renewable resource management. In your words define what that means and how it can be used in a changing and uncertain environment both with respect to water supplies and demands. Over what space and timescales is it applicable, and how can one decide whether or not some plan or management policy will be sustainable? How does this concept relate to the adaptive management concept?

1.11. Identify and discuss briefly some of the major issues and challenges facing water managers today.

Water Resource Systems Modeling: Its Role in Planning and Management

Planning, designing, and managing water resource systems today inevitably involve impact prediction. Impact prediction can be aided by the use of models. While acknowledging the increasingly important role of modeling in water resource 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, i.e., their predictions. As useful as they may be, the results of any quantitative analysis are always only a part of the information that should be considered by those involved in the overall planning and management decision-making process.

2.1 Introduction

Modeling provides a way, perhaps the principal way, of predicting the behavior or performance of proposed system infrastructure designs or management policies. The past 50 years have witnessed major advances in our abilities to model the engineering, economic, ecologic, hydrologic, and sometimes even the institutional or political aspects of large complex multipurpose water resource systems. Applications of models to real

systems have improved our understanding of such systems, and hence have often contributed to improved system design, management, and operation. They have also taught us how limited our modeling skills remain.

When design and management decisions are made, they are based on what the decision-makers assume will take place as a result of their decisions. These predictions are based on qualitative information and beliefs in peoples' heads, as illustrated in Fig. 2.1, possibly informed by quantitative information provided by mathematical or computer-based models as illustrated in Fig. 2.2. Computer-based modeling is used to enhance mental models. These quantitative mathematical models are often 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 behavior of any proposed water resources system design or management policy before it is implemented.

Water resource systems are typically far more complex than what analysts can model and simulate. The reason is not primarily due to computational limitations but rather it is because we do not understand sufficiently the multiple interdependent physical, biochemical, ecological, social, legal, and political (human) processes that govern the behavior of such water resource systems. People and their institutions impact the performance of such systems, and the

Fig. 2.1 Using mental models for prediction

performance of these systems impacts people. System performance is affected by uncertainties in things we can measure and processes we can predict. They are also affected by the unpredictable actions of individuals and institutions as they manage and use water in response to a multitude of impacts they experience in their physical and social environment. Some of these impacts are water related. Others have nothing directly to do with water.

The development and application of models, i.e., the art, science, and practice of modeling, as will be discussed in the following chapters, should be preceded by a recognition of what can and cannot be achieved from the use of models. Models of real-world systems are always simplified representations of those systems. What features of the actual system are represented in a

model, and what features are not, will depend in part on what the modeler 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 modeler, the time and money available, and, perhaps most importantly, the modeler's understanding of the real system and decision-making process.

Developing models is an art. It requires knowledge of the system being modeled, 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 understandings and eventually lead to a good decision clearly requires not only modeling skills but also the

ability to communicate and effectively work with stakeholders and decision-makers.

Models produce information. They do not make decisions or replace those individuals that do. But they can inform them. Water resource planners and managers must accept the fact that decisions may not be influenced by the results of their planning and management models. If model results are not available when needed, they are likely to be ignored when they become available. If model results do not support the preferences of decision-makers, they may also not be considered. 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. And it is unlikely everyone, even so-called experts, will agree on any recommended course of action. Managers or operators may know that not everyone may 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 1998). There is a strong feeling in many cultures and legal systems that fail to act (nonfeasance) is considered 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 risks of failure should not deter us from addressing sensitive planning or policy issues in our models. Modeling 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 resource planners and managers are not intended to be, and rarely are (if ever), a replacement of their judgment. This we have learned, if nothing else, in our over 50 years of modeling experience.

This brief chapter serves to introduce this art of modeling and its applications. The emphasis throughout this book is on application. This chapter is about modeling 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 five decades or more in applying various modeling approaches or tools to real problems with real clients (also see, for example, Austin 1986; Brown et al. 2015; Cai et al. 2013; Castelletti and Soncini-Sessa 2007; Gass 1990; Kindler 1987, 1988; Loucks et al. 1985; Reynolds 1987; Rogers and Fiering 1986; Russell and Baumann 2009; Watkins 2013).

In attempting to understand how modeling can better support planners and managers, it may be useful to examine just what planners and managers of complex water resource systems do. What planners or managers do governs to some extent what they need to know. And what they need to know governs to a large extent what modelers or analysts should be trying to provide. In this book the terms *analysts* or *modelers*, *planners*, and *managers* can be the same person or group of individuals. These terms are used to distinguish the activities of individuals, not necessarily the individuals themselves.
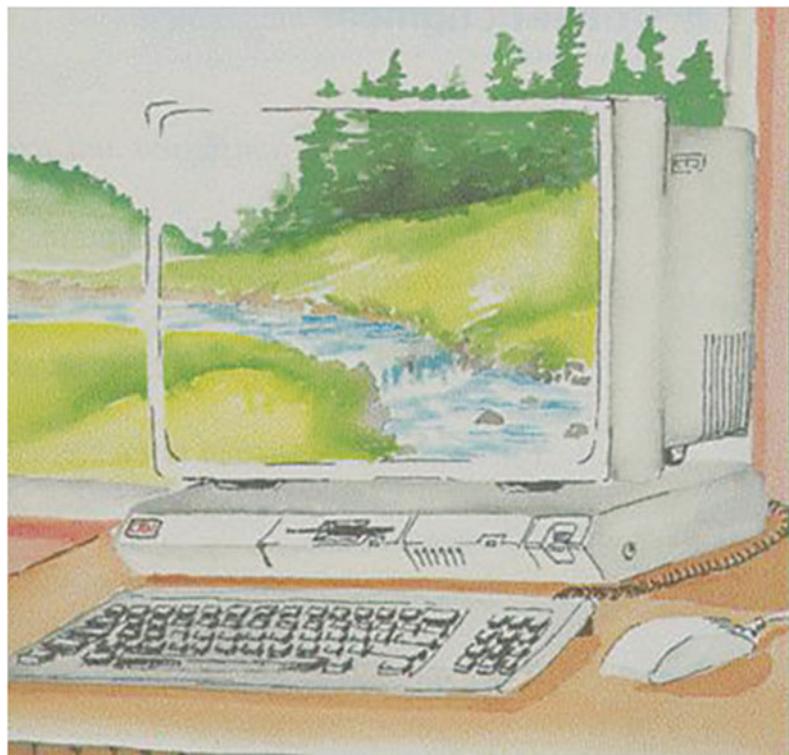
First, we offer some general thoughts on the major challenges facing water resource systems planners and managers, the information they need to meet these challenges, and the role analysts have in helping to provide this information. Next, we review some criteria for evaluating the success of any modeling activity designed to help planners or managers solve real problems. Finally, we argue why we think the practice of modeling is in a continual state of transition, and how current research and development in modeling as well as improvements in computing technology are affecting that transition.

2.2 Modeling Water Resource Systems

As will be discussed in greater detail in the following chapters of this book, there are many types of models and modeling approaches that have been developed and used to identify, study, and evaluate alternative water resource designs, management plans, and operating policies. But before outlining these model types and modeling approaches and how they can be used to best

Fig. 2.2 Using computer-based mathematical models for prediction

meet the needs of planners and managers, it may be useful to describe a specific modeling example based on Borsuk et al. (2001). In this example, a sequence of models are used to assess how effective reductions in upstream nutrient runoff may be in improving the habitat for fish and shellfish in a downstream estuary.

This example is followed by a discussion of the conditions needed that motivate the use of models, whether solely mental (Fig. 2.1) or both mental and mathematical (Fig. 2.2).
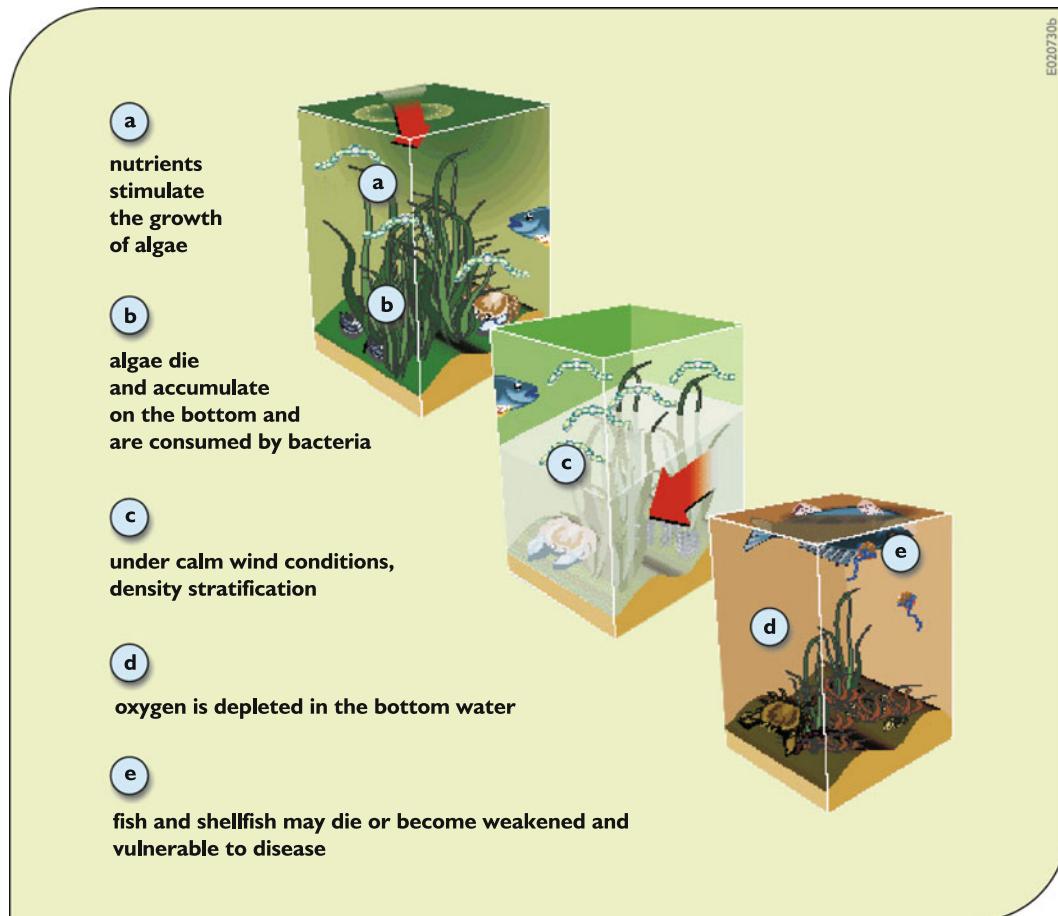
2.2.1 An Example Modeling Approach

Consider for example the sequence or chain of models illustrated in Fig. 2.3 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 economy of the region and the income of many 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 modeling 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 Fig. 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. As a consequence, fish and shellfish may die or become weakened and more vulnerable to disease.

A sequence of models, each providing input data to the next model, can be defined to predict shellfish and fish abundance in the estuary based on upstream nutrient loadings. These models, for

Fig. 2.3 The impacts of excessive nutrients in an estuary
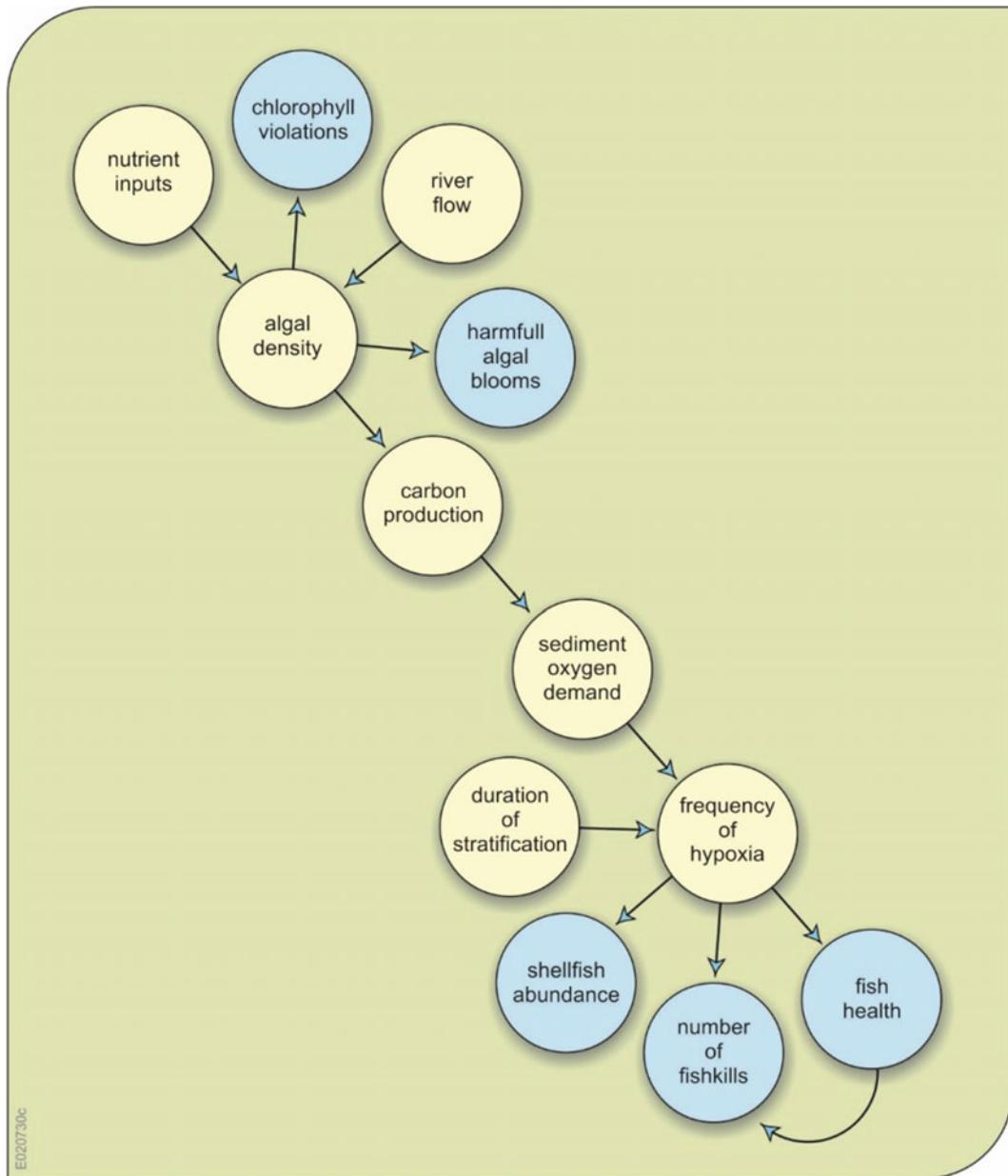
each link shown in Fig. 2.4, can be a mix of judgmental, mechanistic, and/or statistical ones. Statistical models could range from simple regressions to complex artificial neural networks. Any type of model selected will have its advantages and its limitations. Its appropriateness may largely depend on the amount and precision of the data available for model calibration and verification.

The results of any modeling exercise should be expressed in terms meaningful and of interest to those that will be making decisions taking into account those results. In this example 'shell-fish abundance' and 'number of fish-kills' are meaningful indicators to stakeholders and can be related to designated water body use.

2.2.2   Characteristics of Problems to be Modeled

Problems motivating modeling and analyses exhibit a number of common characteristics. These are reviewed here because they provide insight into whether a modeling 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 modeling and other more sophisticated methodologies are likely to be of little use.

Successful applications of modeling are often characterized by:

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

- *A systems focus or orientation*. In such situations attention needs to be devoted to the interdependencies and interactions of elements or components within the system as a whole, as well as to the elements or components themselves.

- *The use of interdisciplinary teams*. In many complex and nontraditional problems, it is not at all clear from the start what mix of 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 modeling 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 modeling methods. Modeling is most likely to be appropriate when:

- The planning and management objectives are reasonably well defined, and organizations and individuals can be identified who can benefit from obtaining and 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 analyzed are describable by reasonably tractable mathematical representations.
- The information needed, such as the hydrological, economic, environmental, and ecological impacts resulting from any decision, can be better estimated through the use of models.
- The values of the model parameters are estimable from readily obtainable data.

2.3   Challenges Involving Modeling

Modeling activities present challenges to those who do it as well as those who sponsor it and may potentially benefit from model results.

2.3.1   Challenges of Planners and Managers

Planners and managers of water resource systems are responsible for solving particular water-related problems or meeting special water resource needs. When they fail, they hear about it. The public lets them know. (Example: the lead contamination in the drinking water of Flint, Michigan USA, after a switch in the water source to reduce costs.) What makes their job particularly challenging is that stakeholders often have different needs and expectations. Furthermore, institutions where water resource planners and managers work (or hire consultants to work for them) are like most institutions these days. They must do what they have been asked to do with limited financial and human resources. Their clients include all of us who use water, or at least all of us who are impacted by the decisions they make.

The overall objective of planners, managers, and operators and their institutions is to provide a service, such as reliable and inexpensive supplies of water, assurance of water quality, production of hydropower, protection from floods, provision of commercial navigation and recreational opportunities, 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 available that will ensure a successful solution to any particular set of water resource management problems. Furthermore, everyone who has had any exposure to water resources planning and management knows one cannot design or operate a water resource system without making compromises. These compromises often involve 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 economical, environmental, ecological, and social impacts, it is the decision-makers who have the more difficult job. They must work with and influence everyone who will be affected by any decision.

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. These processes involve leadership and communication among people and institutions. Leadership and communication skills are learned from experience working with people, not sitting alone working 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 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 modelers or analysts. Some include:

1. identifying creative ways of solving problems.
2. finding out what each interest group wants to know in order to reach an understanding of the issues and a consensus on what to do.
3. developing and using models and presenting their results so that everyone can reach a common or shared understanding and agreement that is consistent with their individual values.
4. making decisions and implementing 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; and finally;
- sciences, engineering, and economics—the technical aspects of their work.

We mention these technical aspects last not to suggest that they are the least important factor to be considered. We do this to emphasize that they are only among many factors and, probably in the eyes of planners and managers, not the most decisive or influential (Ahearne 1988; Carey 1988; Pool 1990; Thissen and Walker 2013; Walker 1987).

So, does the scientific, technical, systematic approach to modeling for planning and management really matter? We believe it can if it addresses the issues of concern to their clients, the planners, and 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 analysts are presenting 'facts' based on the current state of the sciences, sometimes these sciences are not considered relevant. Happily for scientists and engineers, this is not always the case. The challenge of modelers or analysts interested in having an impact on the performance of water resource systems is to become a part of the largely political planning and management process and to contribute towards its improvement.

2.3.2   Challenges of Modelers

To engage in a successful water resource systems study, the modeler must possess not only the requisite mathematical and systems modeling 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 would affect the quantity and quality of the surface and subsurface runoff and how this will impact the quantity and quality of surface waters and ground waters and their ecosystems. These impacts, in turn, might affect the planned development itself, or others downstream. To do this the analysts must have an understanding of the biological, chemical, and 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, hydrologic, 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 knowledge of these disciplines are useful during the comprehensive planning of water resource systems.

Some of the early water resource systems studies were often undertaken with a naive view of the appropriate role and impact of models and modelers in the policymaking process. Policymakers could foresee the need to make a decision. They would ask the systems group to study the problem. These analysts would then model the problem, identify feasible solutions and their consequences, and recommend one or at most a few alternative solutions. The policymakers, after waiting patiently for these recommendations, would then make a yes or no decision. Experience to date suggests the following:

1. A final solution to a water resources planning problem rarely exists; plans and policies 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.
2. For every major decision there are many minor decisions, made by different agencies or management organizations responsible for different aspects of a system.

3. The times normally available to study particular water resources problems are shorter than the times needed to do a thorough study, or if there is sufficient time, the objectives of the original study will likely have significantly shifted by the time the study is completed.

This experience emphasizes some of the limitations and difficulties that any water resource systems study may encounter, but more importantly, it underscores the need for constant communication among the analysts, system planners, managers and operators, and policymakers. 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 can benefit from 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 achieve. Close communication among analysts and all interested stakeholders and decision-makers throughout the modeling 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 often differ from the objectives as understood at the end of a study.

Furthermore, it is helpful if those who will use models, and present the information derived from models to those responsible for making decisions, are intimately involved with model development, solution, and analysis. Only then can they appreciate the assumptions upon which any particular model output is based, and hence adequately evaluate the reliability of the results.

Any water resource systems study that involves only outside consultants, and minimal communication between consultants and planners within a responsible management agency or involved stakeholders, is not likely to have a significant impact on the planning process. Models that are useful tend to be those that are constantly being modified and applied by those involved in plan preparation, evaluation, and implementation.

2.3.3 Challenges of Applying Models in Practice

The clients of modelers or analysts are typically those 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 given what they do, and who will care and how much. The aim of analysts is to provide them with meaningful (understandable), useful, accurate, and timely information. This information is to help them better understand their system, its problems, and alternative ways to address them. In short, the purpose of water resource systems planning and management modeling is to provide useful and timely information to those involved in managing such systems.

Modeling is a process or procedure intended to focus and force clearer thinking and to promote better 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 can benefit from it.

The focus of most books and articles on water resource systems modeling is on modeling 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 four decades ago (Majoni and Quade 1980; Tomlison 1980; Miser 1980; Stokey and Zeckhauser 1977).

There is always a gap between what researchers in water resource systems modeling 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 resource 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 what is worth adapting from this research to meet their needs. Evaluation, testing, and inevitable modifications take time. Not all published research is ready or suited for implementation. By definition research is a work in progress.

How can modelers help reduce the time it takes for new ideas and approaches to be adopted and used in practice? Clearly, practitioners are not likely to accept a new modeling approach or even modeling 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 models or programs. 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 modeling approach.

2.3.4   Evaluating Modeling Success

There are a number of ways one can judge success (or failure) in applying models 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); and
3. How the information derived from the model and its application affected the system design or operation and the lives of those who used the system (outcome success).

The extent to which the models and methods and style of presentation are appropriate for the problem being addressed, the resources and time available for the study, and the institutional environment of the client, are often hard to judge. Publishing in peer-review journals and review panels are two ways of judging. No model or method is without its limitations. Two other obvious indications are the feeling analysts have about their own work and, very importantly, the feeling the clients have about the analysts' work. Client satisfaction may not be an appropriate indicator if, for example, they 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 modeling.

Application or implementation success implies that the methods and/or results developed in the study were seriously considered by those involved in the planning and management process. One should not, it seems to us, judge success or failure based on whether or not any of the model results, i.e., the computer 'printout,' were directly implemented. What one hopes for is that the information and understanding resulting from model application helped define and focus the problem and possible solutions, and helped 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 modeling study will have achieved application or implementation success.

Outcome success is based on what happened to the problem situation once a decision (that was largely influenced by the results of modeling) was made and implemented. The extent to which the information and understanding resulting from modeling 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 based on any of the last two of the three criteria will be strongly dependent on the success of the preceding criteria. Modeling applications may be judged successful based on the first two measures, but perhaps because of unpredicted events, the problems being addressed have become worse rather than improved, or while those particular problems were eliminated, their elimination caused one or more even more severe problems. All of us can think of examples where this has happened. The previously mentioned lead contamination in the drinking water of Flint, Michigan, resulting from trying to reduce costs is one example. Any river restoration project involving the removal of engineering infrastructure is another example of changing objectives or new knowledge following previous decisions that no longer work very well. Who knows—a broader systems study might have helped planners, managers, and decision-makers foresee such consequences, but one cannot count on that. Hindsight is always clearer than foresight. Much of what takes place in the world is completely unpredictable. Given this, it is not clear whether we should hold modelers or analysts, or even planners or managers, completely responsible for any lack of 'outcome success' if unforeseen events change society's goals, priorities, and understanding.

Problem situations and criteria for judging the extent of success are likely to change over time. By the time one can evaluate outcome success, 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   analyzed   and   implemented

modeling 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.

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 have involved a considerable commitment of time and money for both the analysts as well as 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. These interactive interface developments have had a major impact on the state of the practice in using models in the processes of water resources planning and management.

2.4 Developments in Modeling

2.4.1 Technology

The increasing developments in computer technology—from mobile devices to microcomputers and workstations to supercomputers—and all their software applications—have motivated the concurrent development of an impressive set of new models and accompanying software. This software is aimed at facilitating model use and, more importantly, interaction and communication between the analysts or modelers and their clients. This new software includes

1. Interactive approaches to model operation that put users more in control of their computers, models, and data;
2. Computer graphics that facilitate data input, editing, display, and comprehension;

3. Geographic information systems that provide improved spatial analysis and display capabilities;
4. 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;
5. Cloud computing, electronic mail, and the Internet that lets 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;
6. Multimedia systems that permit the use of sound and video animation in analyses, all aimed at improved communication and understanding.

These and other software developments are giving planners and managers improved opportunities for increasing their understanding of their water resource systems. Such developments in technology should continue to aid all of us in converting model output data to information, i.e., it should provide us with a clearer knowledge and understanding of the alternatives, issues, and impacts associated with potential solutions to water resource systems problems. But once again, this improved information and understanding will only be a part of what 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-color computer displays just because we are easily able to, rather than being motivated by the hope that such methods will add to improved understanding of how to solve problems more effectively? Will we provide the safeguards needed to ensure the proper 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 modelers 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 resource systems. We must study their systems and their problems, and we must identify their information needs. We must develop better tools that they themselves 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—you name it—as appropriate. Viewing a water resource system from a single-discipline perspective is rarely sufficient to meet today's water resource systems planning challenges.

Even if we have successfully incorporated all relevant disciplines and data in our analyses, we should have a healthy skepticism 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 or mathematical models), we must admit that many of our assumptions, e.g., parameter values, cannot even be calibrated let alone verified. Our conclusions or estimates can be very sensitive to our assumptions. One of our major challenges is to communicate this uncertainty in understandable ways to those who ask for our predictions.

2.4.2 Algorithms

Accompanying the improvements in the technology of computing that has had an enormous impact on the capability of analysts to address
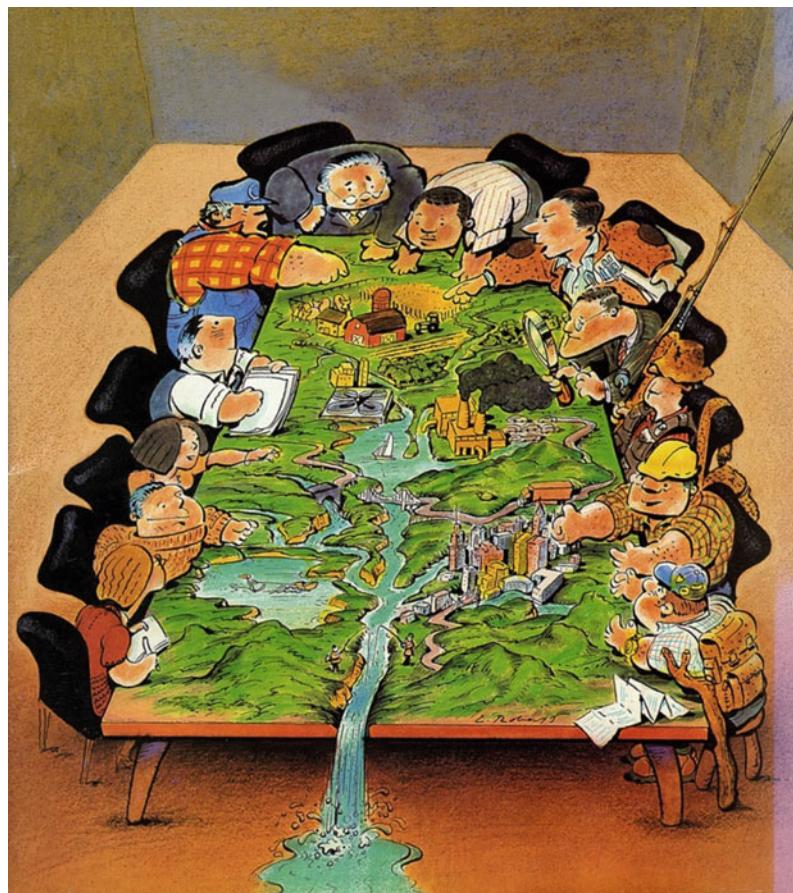
and study increasingly complex issues in water resource systems planning and management, improvements made in the mathematical and computational algorithms have permitted the modeling of more complex systems problems. All our algorithms that have been applied to the analysis of water resource systems, have their strengths and limitations. We still lack the 'perfect' all-purpose algorithm. And it is not likely that we will find one in the future. Probably the major determinant of a particular algorithm or software package chosen to address a particular problem or development opportunity is that which the analyst is most familiar with and experienced in using.

Nevertheless, the menu of available algorithms that can be used for analyses is considerably larger today than what it was when the seminal book on the design of water resource systems (Maas et al. 1962) was published over six decades ago. At that time mathematical programming (constrained optimization) software applied to mainly deterministic linear and nonlinear problems dominated the interests of those working toward improved models for preliminary screening of water resource systems prior to more detailed simulation modeling. Simulations were based on software and constrained by the internal and magnetic tape memory capacity of computers available at that time. Today our focus is more on methods suited for enhancing stakeholder participation. Much of it based on the results of research in artificial intelligence, examples including evolutionary search methods based on biological processes, multi-agent modeling, artificial neural networks, and data mining methods.

2.4.3 Interactive Model-Building Environments

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, such as those shown in Fig. 2.5, who

Fig. 2.5 Stakeholders
involved in river basin
planning and management,
each having different goals
and information needs

typically have different interests, objectives and
opinions about how their water resource 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—vi-
sion? 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 modeling 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 contribute to the solution
of, water resources problems, 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 information we
analysts generate with all our models and com-
puter technology must be understandable, credi-
ble, and timely. It must be just what is needed
when it is needed. It must be not too little and not
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—i.e., 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 that information results in opening up another window of opportunity.
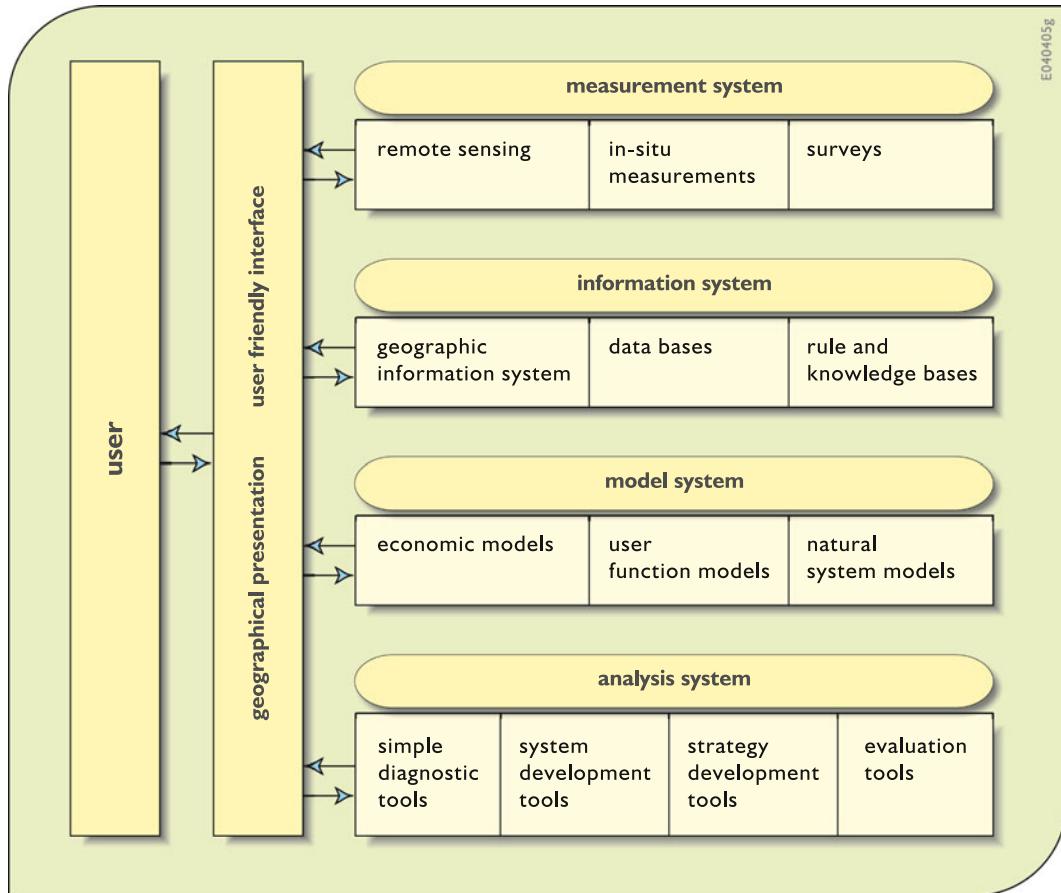
If there is truth in the expression "decision makers don't know what they want until they know what they can get," how do modelers know what decision-makers will need before even they do? How will modelers know what is the right amount and detail of information? How will they know especially if they are to have that information available, and in the proper form, before or at, the time it is needed? Obviously modelers cannot know this. However, over the past three 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 modeling 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. This approach is commonly termed collaborative modeling.

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. 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 when there is a high risk of flooding in Rotterdam harbor is an example of this.

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 do. If they are involved in model building, they will know the assumptions built into their model. Being involved in a joint modeling 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, 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. Though such model-building exercises, it is just possible those involved will gain not only a better understanding of everyone's concerns, but also a common or 'shared' vision of at least how their water resource system (as represented by their model, of course) works. 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 US, one of the major advocates of shared vision or collaborative modeling 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

Fig. 2.6 Common components of many decision support systems

systems (Hamlet et al. 1996a, b, c; Palmer et al. 1995; Werick et al. 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 become a trusted, 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 problems solvers that resembles the natural integration of the conditions they study.

Now the question is how to get all the stakeholders, many who 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 having various backgrounds and interests. We need better tools for building DSSs, not just better DSSs themselves. We need to develop better modeling 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. Researchers need to focus our attention 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 modeling experiences seem to suggest that there are five steps in the modeling process. First, the information the model is to provide is identified. This includes measures of system performance that are of interest to stakeholders. These system performance measures are defined as functions of the behavior or state of the system being modeled. Next this behavior needs to be modeled so the state of the system associated with any 'external' inputs can be predicted. This requires modeling the physical, chemical, biological, economic, ecological, and social processes that take place, as applicable, in the represented system. Third, 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 modeling process is clearly not going to work for those who are not especially trained or experienced or even interested in these modeling 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, and
- give them the interactive control over data input, editing, model operation and output display that they can understand and need in order to make informed decisions.

The challenge in creating such model-building environments is in making them sufficiently useful and attractive so 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?
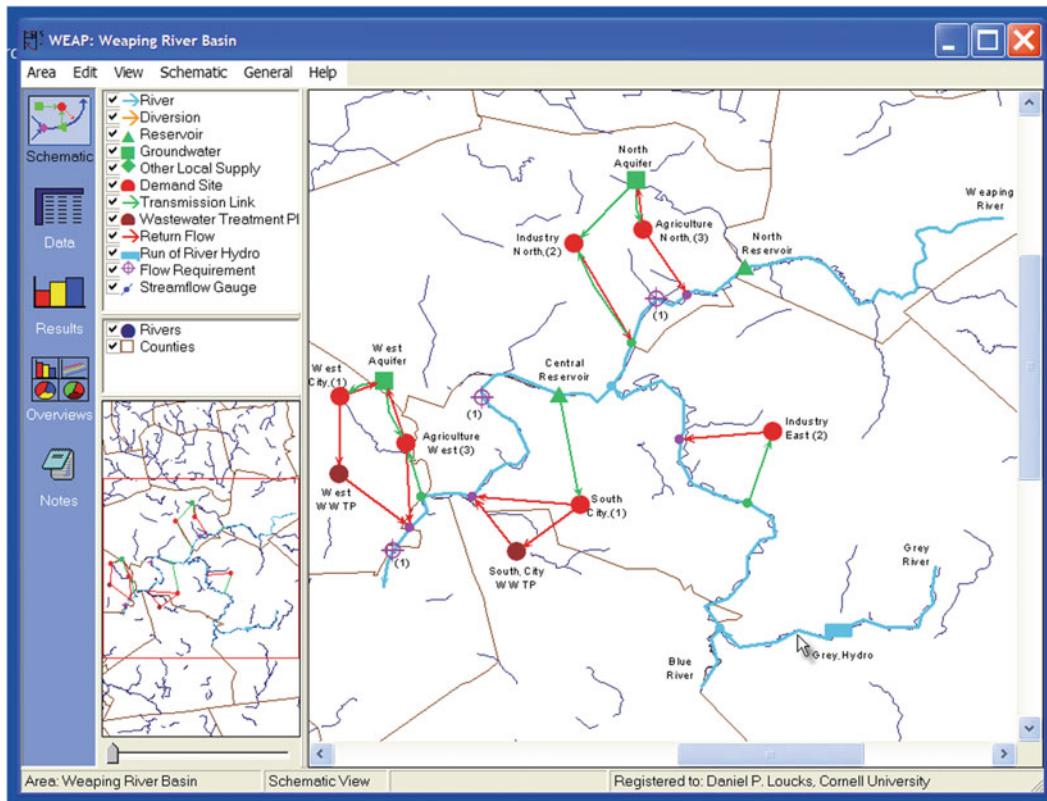
One approach is to develop interactive modeling 'shells' specifically suited to modeling environmental problems. Modeling 'shells' are data-driven programs that become models once sufficient data have been entered into them.

There are a number of such generic modeling shells for simulating water resource systems. AQUATOOL, RIBASIM, MIKE-BASIN and WEAP 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 (e.g., Fig 2.7 from WEAP). Each node and link require data, and these data depend on what that node and link represent, 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 resulting from reservoir operation and/or water allocation policies, then water quality data need not be entered, even though there is the capability of modeling water quality. If water quality outputs are desired, then the user can choose the desired various water quality constituents. Obviously, the more 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 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. This training pertains to the use of the shell and what it can and cannot do. The developers of such shells have removed the need to worry about data base management, solving systems of equations, developing an interactive interface, preserving mass balances and

Fig. 2.7 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 using a node-link network

continuity of flow, and the like. Any assumptions built into the shell should be readily transparent and acceptable by all before its use in any shared vision exercises.

2.4.4 Open Modeling Systems

The next step in shared-vision modeling will be to create a modeling 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 modeling 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).

2.5 **Conclusions**

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

To expect everyone to communicate effectively and to fully understand one another may be asking too much. As written in the Bible (Genesis; Chapter 11, Verses 1–9) there was 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 got 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 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 modelers 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. Who knows, if we are successful, we may even have another visit from the Lord.

Those who are involved in the development of water resource systems modeling 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 helpful or realistic. What modelers can do, however, is to define and evaluate, in different levels of detail, 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 modeling 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 modeling, such modeling 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 resource system. Models permit an evaluation of the economic and physical 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 resource systems planning problems facing water resources planners and managers. How one models any specific water resource problem depends on (a) the objectives of the analysis;

(b) the data required to evaluate the projects; (c) the time, data, money, and computational facilities available for the analysis; and (d) the modeler's knowledge and skill. Model development is an art, requiring judgment 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 judgment in expressing those components and their interrelationships mathematically in the form of a model. This art is to be introduced in Chap. 3.

References

Ahearne, J. F. (1988). Addressing public concerns in science. *Physics Today, 41*(9), 36–42.

Austin, T. A. (1986). Utilization of models in water resources. *Water Resources Bulletin, 22*(1), 49–56.

Borsuk, M. E., Stow, C. A., Higdon, D., & Reckhow, K. H. (2001). A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones. *Ecological Modelling, 143*, 165–181.

Brown, C. M., Lund, J. R., Cai, X., Reed, P. M., Zagona, E. A., Ostfeld, A., et al. (2015, August). The future of water resources systems analysis: Toward a scientific framework for sustainable water management. *Water Resources Research, 51*(8), 6110–6124.

Cai, X., Vogel, R., & Ranjithan, R., (Eds.). (2013, September/October). Role of systems analysis in watershed management. *Journal of Water Resources Planning and Management, 139*(5), 461–592.

Carey, W. D. (1988). Scientists and sandboxes: Regions of the mind. *American Scientist, 76*, 143–145.

Castelletti, A. & Soncini-Sessa, R. (2007). *Topics on system analysis and integrated water resources management*. Amsterdam: Elsevier Ltd. ISBN: 978-0-08-044967-8.

Fedra, K. (1992). Advanced computer applications. *Options*, December.

Gass, S. I. (1990). Model world: 1-lave model, will travel. *Interfaces, 20*(2), 67–71.

Georgakakos, A. P., & Martin, Q. W. (Eds.). (1996, March). An international review of decision support systems in river basin operation. In *Proceedings of the Fifth Water Resources Operations Management Workshop*. Arlington, VA.: ASCE.

Gijsbers, P. J. A., Moore, R. V., & Tindall, C. I. (2002). *Hamonil T: Towards OMI, an open modelling interface and environment to harmonise European developments in water related simulation software*. The Netherlands, Delft: Delft Hydraulics (internal paper).

Goeller B. (1988). A framework for evaluating success in systems analysis. In H. J. Miser & E. S. Quade (Eds.), *Handbook of systems analysis, craft issues and procedural choices*. Chichester: Wiley.

Hamlet, A., et al. (1996a). *Simulating basinwide alternatives using the ACT-ACF shared vision models*. Mobile District, Mobile, Al.: USACE.

Hamlet, A., et al. (1996b). *A history of shared vision modeling in the ACT-ACF comprehensive study*. Mobile District, Mobile, Al.: USACE.

Hamlet, A., et al. (1996c). *Basic Stella II users manual for the ACT-ACF shared vision models*. Mobile District, Mobile, Al.: USACE.

Kindler, J. (1987). Systems analysis approach to water resources management: State of the art. In *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. In *Fourth IFAC International Symposium on Systems Analysis Applied to Management of Water Resources*. Rabat, Morocco: IFAC. October 11–13.

Loucks, D. P., & da Costa, J. R. (Eds.). (1991). *Decision support systems*. NATO Series G. (Vol. 26). Berlin: Springer.

Loucks, D. P., Stedinger, J. R., & Shamir, U. (1985). Modelling water resources systems: Issues and experiences. *Civil Engineering Systems, 2*, 223–31.

Maas, A., Hufschmidt, M. A., Dorfman, R., Thomas, H. A., Jr., Marglin, S. A., & Fair, G. M. (1962). *Design of water-resource systems: New techniques for relating economic objectives, engineering analysis, and governmental planning*. Cambridge, Mass: Harvard University Press.

Majoni, H., & Quade, E. S. (1980). *Pitfalls of analysis*. New York: Wiley.

Miser, H. J. (1980). Operations research and systems analysis. *Science, 209*, 174–182.

Palmer, R. N., Keys, A. M., & Fisher, S. (1995). Empowering stakeholders through simulation in water resources planning. In *Proceedings of the 20th Conference on Water Resources Planning and Management Division*. New York: ASCE.

Pool, R. (1990). Struggling to do science for society. *Science, 248*, 672–673.

Reynolds, P. J. (1987). Future research needs in systems analysis and to application to water resources management. In *International Symposium on Recent Developments and Perspectives in Systems Analysis in Water Resources Management*. Perugia, Italy: WARREDOC.

Rogers, P. P., & Fiering, M. B. (1986). Use of systems analysis in water management. *Water Resources Research, 22*(9), 146S–158S.

Russell, C. S., & Baumann, D. D. (Eds.). (2009). *The evolution of water resource planning and decision making*. Northampton, MA, USA: Edward Elgar.

Shapiro, H. T. (1990). The willingness to risk failure. *Science, 250*(4981), 609.

Simon, H. A. (1998). Prediction and prescription in system modelling. In *15th Anniversary of IIASA, International Institute for Applied Systems Analysis*. Laxenburg, Austria: IIASA.

Stokey, E., & Zeckhauser, R. (1977). *A primer for policy analysis*. New York, W.W: Norton.

Thissen, W. A. H., & Walker, W. E. (Eds.). (2013). *Public policy analysis: New developments* (286 pp). New York, Heidelberg, Dordrecht, London: Springer. doi:10.1007/978-1-4614-4602-6 (ISBN 978-1-4614-4601-9; ISBN 978-1-4614-4602-6 (eBook)).

Tomlison, R. (1980). Doing something about the future. *Journal of the Operational Research Society, 31*, 467–476.

Walker, W. E. (1987). *Important hard problems in public policy analysis (Report P-7282)*. Santa Monica, California: Rand Corporation.

Watkins, D. W. Jr., (Ed.). (2013). Water resources systems analysis through case studies: Data and models for decision making (168 pp). Reston, VA: ASCE Press.

Werick, W. J., Whipple, W. Jr., & Lund, J. (1996). *ACTACF basinwide study*. Mobile District, Mobile, Al.: USACE.

Whelan, G., & Nicholson, T. J. (Eds.). (2002). *Proceedings of the Environmental Software Systems Compability and Linkage Workshop (NRC report NUREG/CP-0177)*. March 7–9, 2000. Richland and Washington.: Pacific Northwest National Laboratory. wrm_ch02.qxd 8/31/2005 Page 56.

Additional References (Further Reading)

Andreu, J., Capilla, J., & Sanchis, E. (1991). Aquatool. A computer assisted support system for water resources research management including conjunctive use. In D. P. Loucks & J. R. da Costa (Eds.), *Decision support systems* (pp. 333–355). Berlin: Springer.

Danish Hydraulic Institute (DHI). (1997). *Mikebasin: Operating manual and description*. Hørsholm: Denmark.

Delft Hydraulics. (2004). *RIBASIM River basin simulation program operating manual and description*. The Netherlands, Delft.

Loucks, D. P. (1995, August). Developing and implementing decision support systems: A critique and a challenge. *Water Resources Bulletin, Journal of American Water Resources Association. 31*(4), 571–82.

O'Callaghan, J. R. (Ed.). (1996). *The interaction of economic, ecology and hydrology*. London: Chapman & Hall.

Raskin, P., Sieber, J. & Huber-Lee, A. (2001). *Water evaluation and planning system: User guide for WEAP21*. Boston, Mass: Tellus Institute.

Exercises

2.1 What is a system?

2.2 What is systems analysis?

2.3 What is a mathematical model?

2.4 Why develop and use models?

2.5 What is a decision support system?

2.6 What is shared vision modeling and planning?

2.7 What characteristics of water resources planning or management problems make them suitable for analysis using quantitative systems analysis techniques?

2.8 Identify some specific water resource systems planning problems and for each problem specify in words possible objectives, the unknown decision variables whose values need to be determined, and the constraints or that must be met by any solution of the problem.

2.9 From a review of the recent issues of various journals pertaining to water resources and the appropriate areas of engineering, economics, planning, and operations research, identify those journals that contain articles on water resources systems planning and analysis, and the topics or problems currently being discussed.

2.10 Many water resource systems planning problems involve considerations that are very difficult if not impossible to quantify, and hence they cannot easily be incorporated into any mathematical model for defining and evaluating various alternative solutions. Briefly discuss what value these admittedly incomplete quantitative models may have in the planning process when nonquantifiable aspects are also important. Can you identify some planning problems that have such intangible objectives?

2.11 Define integrated water management and what that entails as distinct from just water management.

2.12 Water resource systems serve many purposes and can satisfy many objectives. What is the difference between purposes and objectives?

2.13 How would you characterize the steps of a planning process aimed at solving a particular problem?

2.14 Suppose you live in an area where the only source of water (at a reasonable cost) is from an aquifer that receives no recharge. Briefly discuss how you might develop a plan for its use over time.

Models for Identifying and Evaluating Alternatives

Water resources systems are characterized by multiple interdependent components that 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 desired goals or objectives. These alternatives are defined by the values of numerous design, management, and operating policy variables. Constrained optimization together with simulation modeling is the primary way we have of identifying the values of the unknown decision variables that will best achieve specified goals and objectives. This chapter introduces optimization and simulation modeling approaches and describes what is involved in developing and applying them to define and evaluate alternative designs and operating policies.

3.1 Introduction

There are typically many different options available to those planning and managing water resource systems. It is not always clear what set of particular design, management, and operating policy decisions will result in the best overall system performance. That is precisely why modeling is done, to estimate the performance associated with any set of decisions and assumptions, and to predict just how well various economic, environmental, ecosystem, and social or political objectives or goals will be met.

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 achieve 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 (Hipel et al. 2015).

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, 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 in 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, such as in the Florida Everglades discussed in Chap. 1, 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, 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 modeling approaches commonly used to study and analyze water resources systems. The modeling 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 modeling methods. This is not a text on the state of the art of modeling. More realistic and more complex problems usually require much bigger and more complex models than those introduced in this book, but these bigger and more complex models are often based on the principles and techniques presented here.

The emphasis here is on the art of model development: just how one goes about constructing a model that will provide information needed to study and address particular problems, 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 example problems 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 modeling approaches and solution methods presented in this book to develop the models needed to analyze their own particular problems.

The water resource planning and management problems and issues used here, or any others that could have been used to illustrate model development, 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 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 to include and what not to include and what parameter values to assume in models of water resource systems requires judgment, experience, and knowledge of the particular problem(s) being addressed, the system being modeled and the decision-making environment. Understanding the contents of, and performing the exercises pertaining to, this chapter will be a first step toward gaining some judgment and experience in model development and solution.

3.1.1 Model Components

Mathematical models typically contain one or more algebraic equations or inequalities. These expressions include variables whose values are assumed to be known and others that are unknown and to be determined. Variables that are assigned known values are usually called *parameters*. Variables having unknown values

that are to be determined by solving the model are called *decision variables*. Models are developed for the primary purpose of identifying the best values of the latter and for determining how sensitive those derived values are to the assumed parameter values.

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 waste removal 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 at specific sites and times, ecological habitat suitability values or deviations from particular ecological, economic, or hydrological targets.

Models describe, in mathematical terms, the system being analyzed 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 its storage capacity. Both the storage volume over time and the reservoir capacity might be unknown and are to be determined. 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 among 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 Chap. 8 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 determine the costs and benefits or the values of other measures of system performance associated with that particular management plan or policy. While the components of optimization and simulation models can include system performance indicators, model parameters and constraints, the process of model development and use also includes people. The drawing shown in Fig. 3.1 (and in Chap. 2 as well) illustrates some interested stakeholders busy studying their river basin, in this case perhaps with the use of a physical simulation model. (Further discussion of stakeholder involvement in the planning and management process is in Chap. 13).

Whether a mathematical model or physical model is being used, one important consideration is that if the modeling exercise is to be of any value, it must provide the information desired and in a form that the interested stakeholders and decision-makers can understand.

3.2   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 choosing the one that best satisfies a particular objective or set of objectives. The processes of plan formulation and selection involve modeling and communication among all interested stakeholders, as the picture in Fig. 3.1 suggests.

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

Fig. 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

3.2.1   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 needed to maintain a desired water quality in a receiving stream, the capacities, and operating rules of multipurpose reservoirs and hydropower plants, and the extent and reliability of floodplain protection derived 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 Chap. 9).

To illustrate this plan formulation process, consider the task of designing a tank that can store a fixed volume, say V, of water. Once the desired shape has been determined, the task is to build a model that can determine the values of all the design variables and the resulting cost. Different designs result in different sizes and amounts of materials, and hence different costs. Assume the purpose of the model is to define the set of design variable values that results in the minimum total cost, for a range of values of the required volume, V.

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 unknown design variables are the tank length, L, width, W, and height, H. 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 its cost. But for now assume V is known.

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. Assume that we know the average costs per unit area of the base, sides, and top of the tank. 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 meters, then the areas will be expressed in units of square meters and the volume will be expressed in units of cubic meters. The average unit costs will be expressed in monetary units per square meter.

The final step of model building is to specify all the relations among the model parameters and decision variables. This includes defining the objective (cost) function (in this case just one unknown variable, *Cost*) and all the conditions that must be satisfied while achieving that objective. It is often helpful 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 convert the word model to 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.

- The total side area is twice the products of length times height and width times height.
- 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 top and base area is the product of length times width.
- The volume of the tank must at least equal the required volume capacity.
- The volume of the tank is the product of the length, width, and height of the tank.

Converting each of the above conditions to mathematical expressions using the notation defined above and inventing new notation when needed results in:

- Total cost equals the sum of the costs of the base, the sides, and the top.
  Cost = sidecost + basecost + topcost
- Cost of the sides is the cost-per-unit area of the sides times the total side area.
  $sidecost = C_{\text{side}} (sidearea)$
- The total side area is twice the products of length times height and width times height.
  sidearea = $2(LH+WH)$
- Cost of the base is the cost-per-unit area of the base times the total base area.
  basecost = C_{base} (basearea)
- Cost of the top is the cost-per-unit area of the top times the total top area.
  topcost = C_{top} (toparea)
- The top and base area is the product of length times width.
  toparea = basearea = LW
- The volume of the tank must at least equal the required volume capacity.
  tankvolume $\geq V$
- The volume of the tank is the product of the length, width, and height of the tank.
  tankvolume = LWH

Combining some of the above conditions, a mathematical optimization model can be written as:

$$\text{Minimize } Cost \qquad (3.1)$$

Subject to:

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

$$LWH \geq V \qquad (3.3)$$

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 the product LWH will equal the required volume 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 V.

In this model, the unknown decision variables include Cost, L, W, and H

The least-cost solution (using methods discussed in the next chapter) is

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

and

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

or

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

The modeling exercise should not end here. If there is any doubt about the value of any of the parameters, a sensitivity analyses can 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 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 or assumptions are important and what are not can guide monitoring and data collection efforts. This is a beneficial attribute of modeling 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 well as cost 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.

In the above discussion, the required volume capacity, V, has been assumed to be known. In reality, it too may be a decision variable, and what would be of greater value to decision-makers is knowing the relationship between various assumed values of V and their respective minimum costs. Such a cost function can be defined by solving the model (defined by Eqs. 3.1, 3.2 and 3.3) for various values of V.

Whatever the final outcome of our modeling 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.

3.2.2 Plan Selection

There are various approaches to finding the "best" plan or best set of decision variable values that satisfy an objective or goal. By trial and

error, one could identify alternative plans, evaluate the performance of each plan, and select the particular plan whose performance is judged better than the others. This process could include a systematic simulation of a range of possible solutions in a search for the best. When there are 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 3.1–3.3 represent an optimization problem. There are an infinite number of feasible tank designs, i.e., alternative values of L, W, and H that satisfy the volume requirement. Our job is to find the least-cost one. We can do this using a mathematical optimization method. 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 analyzed 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 can be 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 Eq. 3.3. For example, the tank has to be able 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 storage capacities. 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 or infrastructure projects. These are a few examples of physical, legal, and financial conditions or constraints that may restrict the ranges of decision 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, Eq. 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 analyzed. They define the system components and their interrelationships, 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 alternatives 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 a model may not be optimal with respect to those involved in a planning or decision-making process. To repeat what was written in Chap. 2, models are used to provide information (useful

information, one hopes), to the decision-making process. Model solutions are not replacements for judgments of individuals involved in the decision-making process.
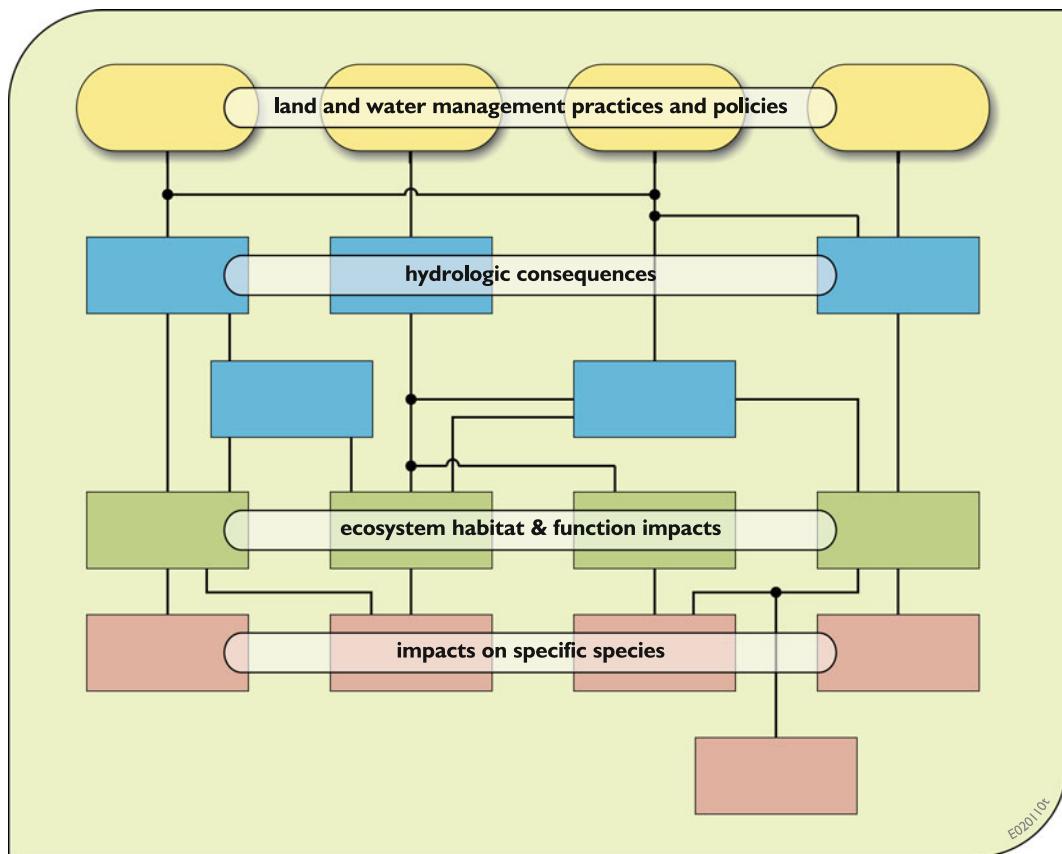
3.3   Conceptual Model Development

Prior to the selection or development of a quantitative model, it is often useful to develop a conceptual one. Conceptual models are non-quantitative representations of a system. The system components and their interactions are defined often by diagrams similar to Fig. 3.2.

Figure 3.2 illustrates the form of a conceptual model. This example conceptual model defines the 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 variables whose values are unknown and can vary, and parameters whose values are assumed known.

The values of the model's parameters need to be determined. Model calibration involves finding the best values for these parameters. Calibration is based on comparisons of the model results with observed data. Optimization methods can sometimes be used to identify the values of model parameters. This is called model calibration or identification. (Illustrations of the use of

Fig. 3.2 An outline of a conceptual model without its detail (i.e., what exactly each component or box represents), showing the links representing interactions among components and between management decisions and specific system impacts

optimization for estimating model parameter values are presented in the following chapter.) Sensitivity analysis may serve to identify the impacts of uncertain parameter values and show which parameter values substantially influence the model's results or solutions. Following calibration, the remaining uncertainties in the model predictions may be quantified in an uncertainty analysis as discussed in Chap. 8.

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 determine whether or not the model describes the system behavior sufficiently accurately.

3.4 Simulation and Optimization

The modeling approach to tank design discussed in the previous section focused on the use of optimization methods to identify the preferred design variable values. Similar optimization methods can be used to identify preferred design variable values and operating policies for urban stormwater runoff control or multiple reservoir systems, given various assumptions regarding parameter values and design and operating objectives. 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 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, environmental, and social 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, 30 infrastructure capacity variables whose values are to be determined. Even if only two possible values are assumed for each of the 30 variables (such as to exist at some predetermined capacity or not), 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— 24 h per day. Most simulation models of water resources systems contain many more variables, each having a larger range of feasible values, and are much more complex than this simple 30-binary-variable example. Mathematically, if not in reality, there could be an infinite combination of feasible values for each of the decision variables.

Simulation works well 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 judgment concerning which alternatives to simulate is often 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 events over a number 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.

3.4.1   Simulating a Simple Water Resources System

Consider the case of a potential reservoir releasing water to downstream users (Fig. 3.3). 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 and more reliable downstream flow conditions will offset the costs of the reservoir.

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 of the released water to allocate to each user and to the river downstream of the users.
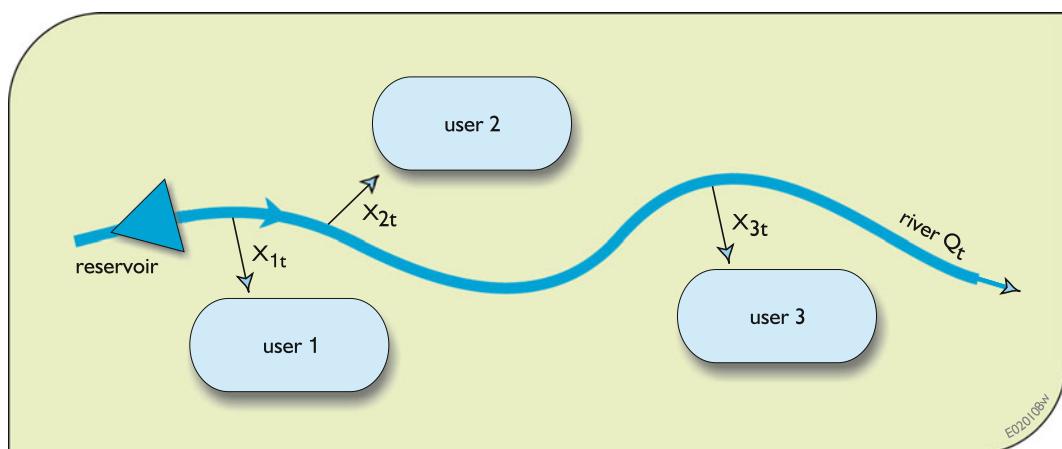
There are literally an infinite number of possible design and operating policy variable values. The next section will address the problem of screening these alternatives to find those values that are most worthy of further study using simulation.

For this simple illustration assume the operating and allocation policies are as shown in Fig. 3.4. Also for simplicity assume they apply to each discrete time period. The reservoir
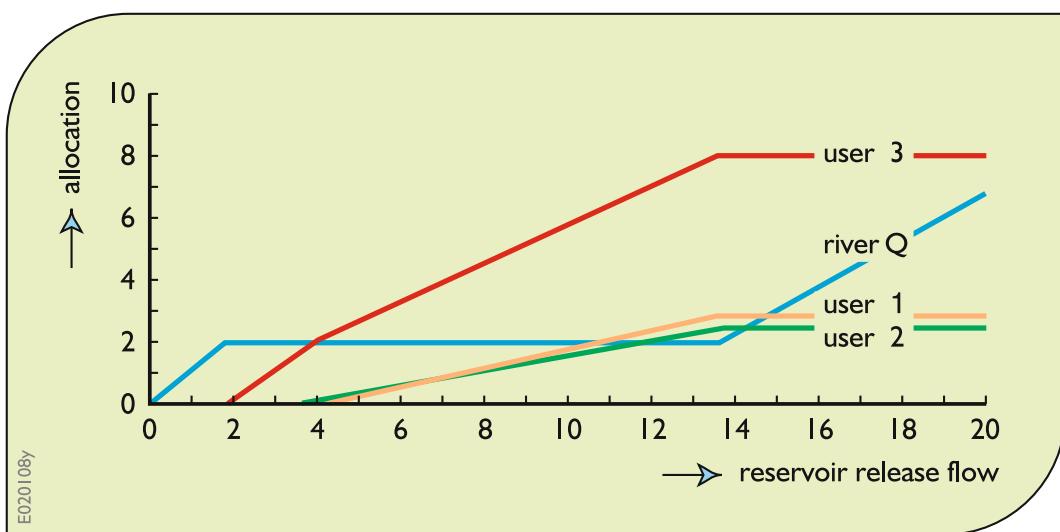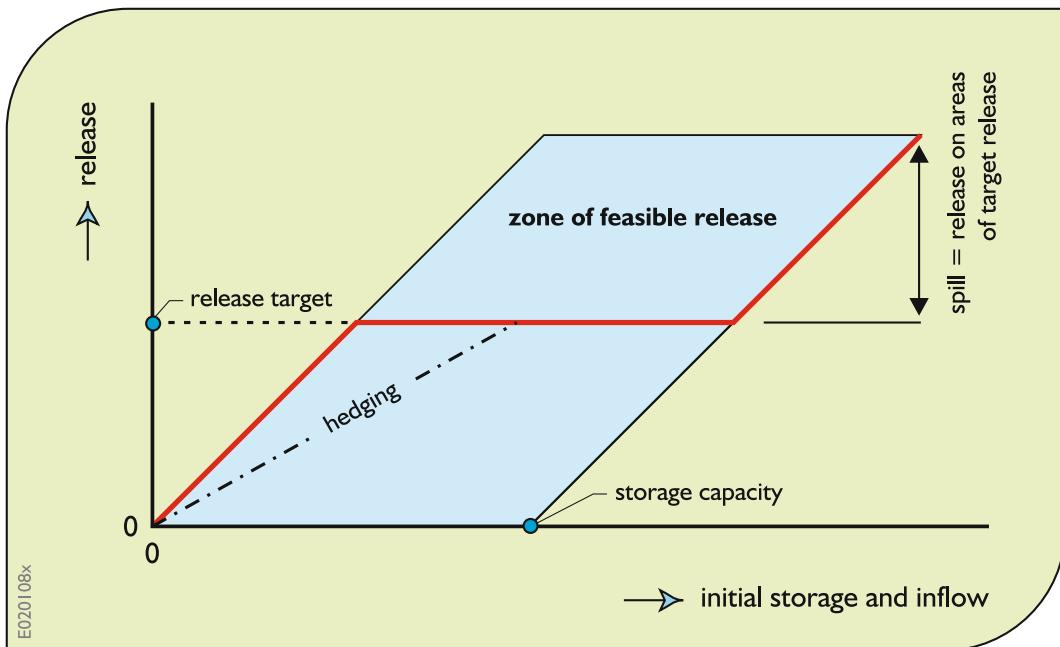
operating policy, shown as a red line in upper Fig. 3.4, attempts to meet a release target. If insufficient water is available, all the water will be released in the time period. If the inflow exceeds the target flow and the reservoir is full, a spill will occur.

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 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 Fig. 3.4. The dash–dot line in Fig. 3.4 is one such hedging function. Once defined, any reservoir operating policy can be simulated.

The simulation process for the three-user system is shown in Fig. 3.5. It 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 based on the release policy (upper Fig. 3.4). 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

**Fig. 3.3** Conceptual model of a reservoir water allocation system to be simulated
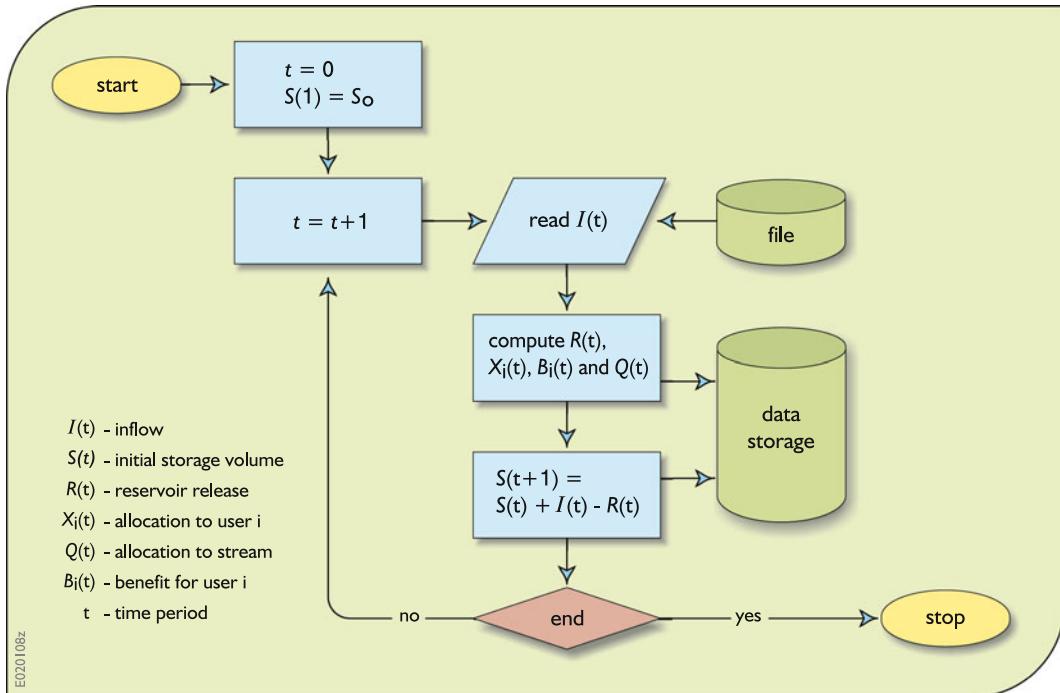
Fig. 3.4 Reservoir operating policy defining the reservoir release to be made as a function of the current storage volume and current inflow and the allocation policy for the river flow downstream of the reservoir. The *blue zone* in the reservoir release policy indicates the zone of feasible releases. It is physically impossible to make releases represented by points outside that *blue zone*

(lower Fig. 3.4). The resulting benefits can be calculated and stored in an output database.

Additional data including storage volumes, releases, and the allocations themselves can also be stored in the output database, as desired. The simulation 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, operation policies and/or allocation policies.

Fig. 3.5 Flow diagram of the reservoir—user allocation system simulation process. The simulation terminates after some predefined number of simulation time steps

It would not be too difficult to write a computer program to perform this simulation. In fact, it can be done on a spreadsheet. However as easy as 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 also include hydrological data and design and operating policy data. These generic simulation programs are capable of simulating surface and ground water flows, storage volumes

and qualities under a variety of system infrastructure designs and operating policies.

3.4.2   Defining What to Simulate

Before the simple system shown in Fig. 3.3 can be simulated the design and operating policy of the system, i.e., the information shown in Fig. 3.4 needs to be defined. One way to do this is to use optimization. Optimization is driven by an objective function. Assume an overall measure of system performance has been decided upon, and can be expressed as a function of the decision variables. These decision variables include all the information in Fig. 3.3, namely the reservoir capacity and reservoir storage and release and water user allocation decisions in each time period. Of interest are the values of these decision variables that achieve the highest level of system performance. The use of an optimization model will help in defining those variable values.

Expressed in words, the optimization model is to be developed and used to identify the decision variable values that maximize system performance. Let $B(K, S, R, A)$ represent the overall system performance measure, as a function of the reservoir capacity K, and all the initial storage volumes, $S(t)$, releases, $R(t)$, and water allocations to users i, $A(i,t)$, in each time period t for a total of T time periods. Hence the objective is to

$$\text{maximize } B(K, S, R, A) \qquad (3.7)$$

while making sure that a mass balance of water is maintained in the reservoir over time.

$$S(t) + \text{Inflow}(t) - R(t) = S(t+1) \quad \text{for each period } t$$
$$\text{(and period } T + 1 = 1)$$
$$(3.8)$$

These mass balance equations define the relationship between initial, $S(t)$, and final, $S(t + 1)$ storage volume values in each period t, and equate the final storage value in each period to the initial value in the following period. Finally, it assumes the entire simulation process repeats itself after every T years.

The next set of constraints ensure that the storage volumes, $S(t)$, do not exceed the reservoir storage capacity K and that the allocations, $A(i, t)$, to the three water users i do not exceed the reservoir release, $R(t)$, less the amount to remain in the stream, $Q(t)$.

$$S(t) \leq K \quad \text{for each period } t. \qquad (3.9)$$

$$A(1,t) + A(2,t) + A(3,t) \leq R(t) - Q(t) \quad \text{for each period } t. \qquad (3.10)$$

This simple example ignores many of the details one should consider when modeling reservoirs and water users, and many of these details will be discussed, and modeled, in subsequent chapters. But for now the model is sufficient to find values for each decision variable shown in upper portion of Fig. 3.4. The allocation policies shown in the lower portion of Fig. 3.4 can be obtained by solving a separate single-period optimization model containing only the allocation benefits as the objective, $B(A)$, and constraint 3.10 for a single period, and various values of the water available, $R - Q$, assuming the benefits, $B(A)$, do not change over time.

$$\text{Maximize } B(A(1), A(2), A(3)) \qquad (3.11)$$

Subject to:
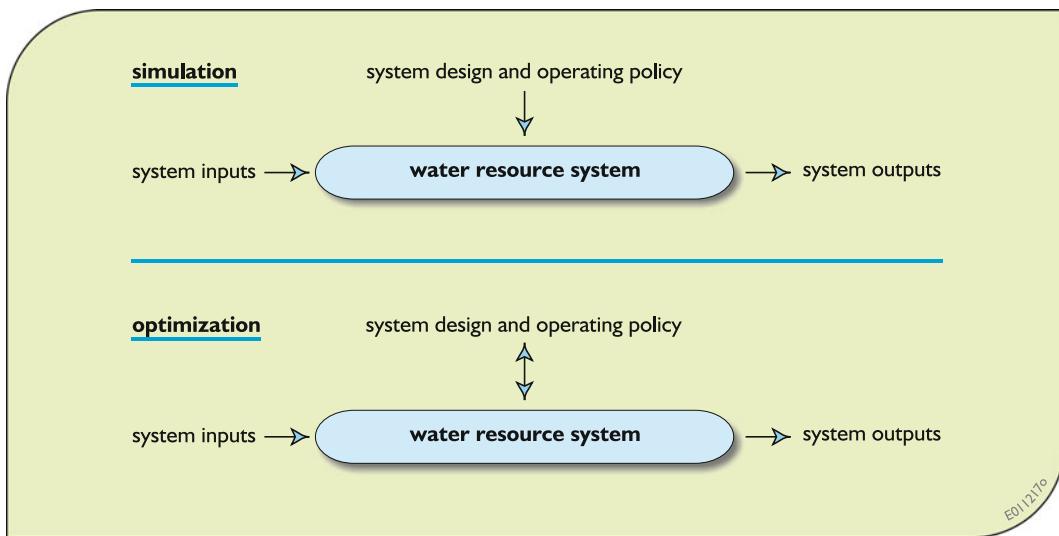
$$A(1) + A(2) + A(3) \leq R - Q \quad \text{for various values of } R, \text{ given } Q. \qquad (3.12)$$

3.4.3 Simulation Versus Optimization

Unlike simulation models, the solutions of optimization models are based on objective functions that are to be maximized or minimized. The objective function and 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 (Eqs. 3.1, 3.2 and 3.3) is an example of an optimization model. So is the reservoir water allocation model, Eqs. 3.7–3.10 and the single-period allocation model Eqs. 3.11 and 3.12.

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. At best these assumptions can only approximate reality.

The assumptions made to permit model solution by optimization solution procedures (algorithms), may justify a more detailed and more realistic simulation to check and improve on any solution obtained from that optimization. While the results from a simulation model may be more realistic, both optimization and simulation models are approximations of the real system being modeled. The optimal solution of any model is optimal *only with respect to the particular model*,

Fig. 3.6 Distinguishing between simulation and optimization modeling. Simulation addresses "what if" questions; optimization can address "what should be" questions. Both types of models are typically used in water resources planning and management studies

not necessarily with respect to the real system. It is important to realize this limited meaning of the word "optimal," a term commonly found in papers published by water resources and other systems analysts, planners, and engineers.

Figure 3.6 illustrates the broad differences between simulation and optimization. Optimization models need explicit expressions of objectives. Simulation models do not. Simulation simply addresses "what-if" scenarios—what may happen if a particular scenario is assumed or if a particular decision is made. Users of simulation models 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 provide more precise estimates of the impacts that may result from those decisions.

While optimization will tell us 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 more detailed simulations. 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   Conclusions

This chapter has reviewed some basic types of models and presented guidelines for their use. Generic models for water resources system analyses are increasingly becoming available, saving many organizations 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 analyzing and presenting their results. The information provided in this book is intended to help those who wish to build their modeling skills. Such skills will be useful to those involved in water resource systems planning and management activities. Such skills may be useful even to those who are expected to oversee or evaluate the model results of others (say from various UN, World Bank, or national aid agencies) who are involved in analyzing particular water resource systems in particular regions of the world.

Reference

Hipel, K. W., Fang, L., Cullmann, J., & Bristow, M. (Eds.). (2015). *Conflict resolution in water resources and environmental management*. Cham, Heidelberg, New York, Dordrecht, London: Springer.

Additional References (Further Reading)

Abbott, M. B. (1999). Introducing hydroinformatics. *Journal of Hydroinformatics, 1*(1), 3–20.

Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water Resources Research, 23*, 1393–1442.

Biswas, A. K. (Ed.). (1997). *Water resources: Environmental planning, management, and development*. New York: McGraw-Hill.

Bobba, A., Ghosh, S., Vijay, P., & Bengtsson, L. (2000). Application of environmental models to different hydrological systems. *Ecological Modelling, 125*, 15–49.

Booch, G. (1994). *Object-oriented analysis with applications*. Benjamin Cummings: Redwood City.

Chan, H. W., & Lam, P. F. (1997). Visualizing Input and output analysis for simulation. *Simulation Practice and Theory, 5*(5), 425–453.

Edsall, R. M., Harrower, M., & Mennis, J. L. (2000). Tools for visualizing properties of spatial and temporal periodicities in geographic data. *Computers & Geosciences, 26*(1), 109–118.

Fedra, K., & Jamieson, D. G. (1996). An object-oriented approach to model integration: A river basin information system example. In: K. Kowar & H. P. Nachtnebel (Eds.), *Proceedings of the HydroGIS'96 Conference,* (vol. 235, pp. 669–676). Vienna: IAHS Press.

Guariso, G., & Werthner, H. (1989). *Environmental decision support systems*. New York: Ellis Horwood-Wiley.

Ho, J. K. K., & Sculli, D. (1997). The scientific approach to problem solving and decision support systems. *International Journal of Production Economics, 48* (3), 249–257.

Hufschmidt, M. M., & Fiering, M. B. (1966). *Simulation in the design of water resource systems*. Cambridge, Massachusetts: Harvard University Press.

Lam, D. C. L., Mayfield, C. I., & Sawyne, D. A. (1994). A prototype information system for watershed management and planning. *Journal of Biological Systems, 2*(4), 499–517.

Lam, D., & Swayne, D. (2001). Issues of EIS software design: Some lessons learned in the past decade. *Environmental Modeling and Software, 16*(5), 419–425.

LAW, A. M., & KELTON, W. D. (2000). *Simulation modeling and analysis* (3rd ed.). Boston, Mass: McGraw-Hill.

Loucks, D. P., Stedinger, J. S., & 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., & Fair, G. M. (1962). *Design of water resources systems*. Cambridge, UK: Harvard University Press.

Reitsma, R. F., & Carron, J. C. (1997). Object-oriented simulation and evaluation of river basin operations. *Journal of Geographic Information and Decision Analysis, 1*(1), 9–24.

Rizzoli, A. E., Davis, J. R., & Abel, D. J. (1998). Model and data integration and re-use in environmental decision support systems. *Decision Support Systems, 24*(2), 127–144.

Rizzoli, A. E., & Young, W. J. (1997). Delivering environmental decision support systems: Software tools and techniques. *Environmental Modeling & Software, 12*(2–3), 237–249.

Sargent, R. G. (1982). Verification and validation of simulation models. In F. E. Cellier (Ed.), *Progress in modelling and simulation* (pp. 159–169). London: Academic Press.

Sargent, R. G. (1984a). A tutorial on verification and validation of simulation models. In: S. Sheppard, U. Pooch, & 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, & M. S. Elzas (Eds.), *Simulation and model-based methodologies: An integrative view* (pp. 537–55). Berlin: Springer (No. 10 in the series: NATO ASI Series F: Computer and Systems Sciences).

Scholten, H., Van Waveren, R. H., Groot, S., Van Geer, F. C., Wösten, J. H. M., Koeze, R. D., et al. (2000). *Good Modelling Practice in Water Management, Proceedings Hydroinformatics 2000*. Cedar Rapids, Iowa: International Association for Hydraulic Research.

Singh, V. P. (Ed.). (1995). *Computer models of watershed hydrology*. Littleton, Colorado: Water Resources Publications.

Sun, L., & Liu, K. (2001). A method for interactive articulation of information requirements for strategic decision support. *Information and Software Technology, 43*(4), 247–263.

Van Waveren, R. H., Groot, S., Scholten, H., Van Geer, F. C., Wösten, J. H. M., Koeze, R. D., et al. (1999). *Good modelling practice handbook* (STOWA report 99–05). Lelystad, The Netherlands: Dutch Department of Public Works, Institute for Inland Water Management and Waste Water Treatment (report 99.036).

Watkins, D. W. Jr., (Ed.). (2013). Water resources systems analysis through case studies: Data and models for decision making (168 pp). Reston, VA: ASCE Press.

Young, P. (1998). Data-based mechanistic modeling of environmental, ecological, economic and engineering systems. *Environmental Modeling & Software, 13*, 105–122.

Young, W. J., Lam, D. C. L., Ressel, V., & Wong, J. W. (2000). Development of an environmental flows decision support system. *Environmental Modeling & Software, 15*, 257–265.

Zeigler, B. P. (1976). *Theory of modelling and simulation*. Malabar, Florida: Robert E. Krieger.

Zeiler, M. (1999). *Modeling our world: The ESRI guide to geodatabase design*. Redlands, California: ESRI.

Exercises
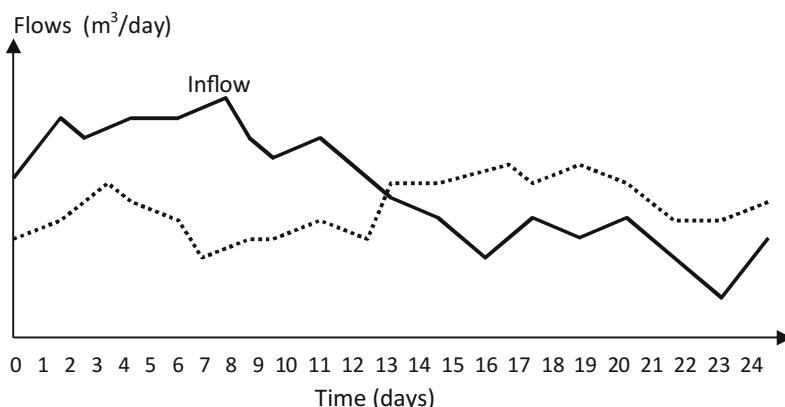
3.1 Briefly outline why multiple disciplines are needed to efficiently and effectively manage water resources in major river basins, or even in local watersheds.

3.2 Describe in a page or two what some of the water management issues are in the region where you live.

3.3 Define adaptive management, shared vision modeling, and sustainability.

3.4 Distinguish what a manager does from what an analyst (modeler) does.

3.5 Identify some typical or common water resources planning or management problems that are suitable for analysis using quantitative systems analysis techniques.

3.6 Consider the following five alternatives for the production of energy (10^3 kwh/day) and irrigation supplies (10^6 m^3/month):

Alternative	Energy production	Irrigation supply
A	22	20
B	10	35
C	20	32
D	12	21
E	6	25

Which alternative would be the best in your opinion and why? Why might a decision-maker select alternative E even realizing other alternatives exist that can give more hydropower energy and irrigation supply?
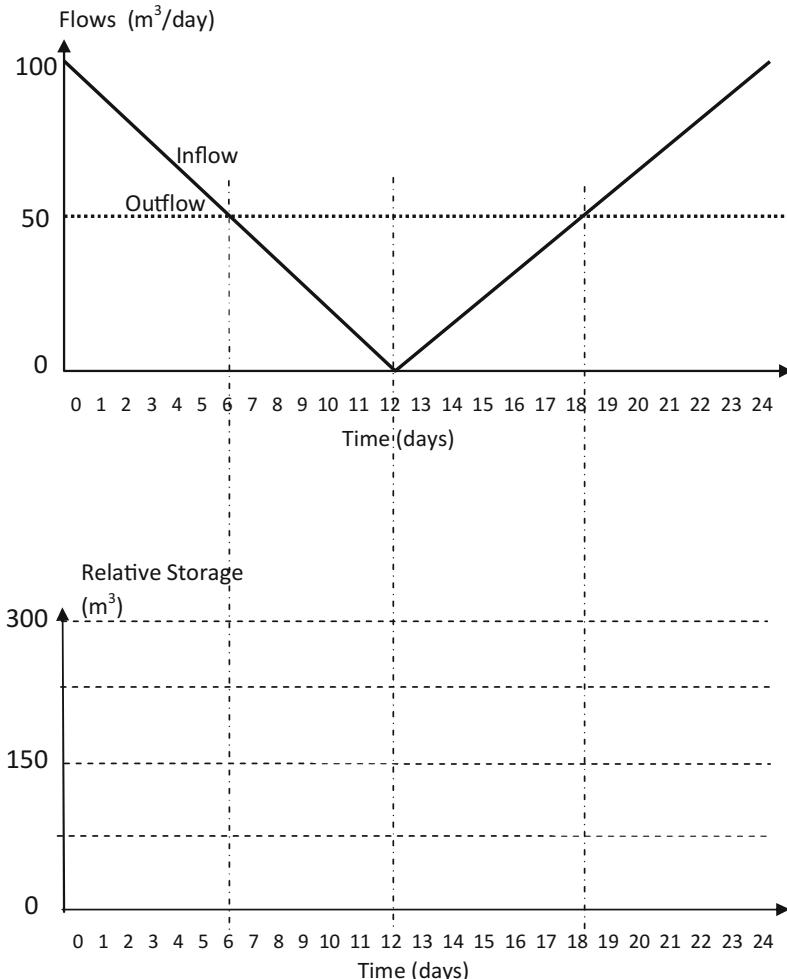
3.7 Define a model similar to Eqs. 3.1–3.3 for finding the dimensions of a cylindrical tank that minimizes the total cost of storing a specified volume of water. What are the unknown decision variables? What are the model parameters? Develop an iterative approach for solving this model.

3.8 Briefly distinguish between simulation and optimization.

3.9 Consider a tank, a lake or reservoir or an aquifer having inflows and outflows as shown in the graph below.

(a) When was the inflow its maximum and minimum values?
(b) When was the outflow its minimum value?
(c) When was the storage volume its maximum value?
(d) When was the storage volume its minimum value?

(e) Write a mass balance equation for the time series of storage volumes assuming constant inflows and outflows during each time period.

3.10 Given the changing inflows and constant outflow from a tank or reservoir, as shown in the graph below, sketch a plot of the
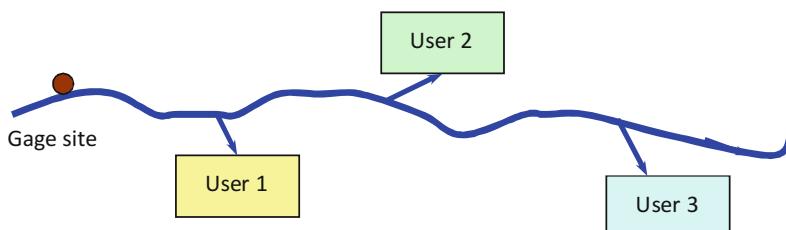
storage volumes over the same period of time, beginning at 150. Show how to determine the value of the slope of the storage volume plot at any time from the inflow and outflow (= 50 m^3/day) graph below.

3.11 Describe, using words and a flow diagram, how you might simulate the operation of a storage reservoir over time. To simulate a reservoir, what data do you need to have or know?

3.12 Identify and discuss a water resources planning situation that illustrates the need for a combined optimization-simulation study in order to identify the best plan and its impacts.

3.13 Write a flow chart/computer simulation program for computing the maximum yield of water that can be obtained given any value of active reservoir storage capacity, K, using.

Year y	Flow Q_y	Year y	Flow Q_y
1	5	9	3
2	7	10	6
3	8	11	8
4	4	12	9
5	3	13	3
6	3	14	4
7	2	15	9
8	1		

Find the values of the storage capacity K required for yields of 2, 3, 3.5, 4, 4.5, and 5.

3.14 How many different simulations of a water resource system would be required to ensure that there is at least a 95% chance that the best solution obtained is within the better 5% of all possible solutions that could be obtained? What assumptions must be made in order for your answer to be valid? Can any statement be made comparing the value of the best solution obtained from the all the simulations to the value of the truly optimal solution?

3.15 Assume in a particular river basin 20 development projects are being proposed. Assume each project has a fixed capacity and operating policy and it is only a question of which of the 20 projects would maximize the net benefits to the region. Assuming 5 min of computer time is required to simulate and evaluate each combination of projects, show that it would require 36 days of computer time even if 99% of the alternative combinations could be discarded using "good judgment." What does this suggest about the use of simulation for regional interdependent multiproject water resources planning?

3.16 Assume you wish to determine the allocation of water X_j to three different users j, who obtain benefits $R_j(X_j)$. The total water available is Q. Write a flow chart showing how you can find the allocation to each user that results in the highest total benefits.

3.17 Consider the allocation problem illustrated below.

The allocation priority in each simulation period t is:

First 10 units of streamflow at the gage remain in the stream.

Next 20 units go to User 3.

Next 60 units are equally shared by Users 1 and 2.

Next 10 units go to User 2.

Remainder goes downstream.

(a) Assume no incremental flow along the stream and no return flow from users. Define the allocation policy at each site.

This will be a graph of the allocation as a function of the flow at the allocation site.

(b) Simulate this allocation policy using any river basin simulation model such as RIBASIM, WEAP, Modsim, or other selected model, including your own, for any specified inflow series ranging from 0 to 130 units.

An Introduction to Optimization Models and Methods

Water resource systems are characterized by multiple interdependent components that together produce multiple economic, environmental, ecological, and social impacts. As discussed in the previous chapter, planners and managers working toward improving the design and performance of these complex systems must identify and evaluate alternative designs and operating policies, comparing their predicted performance with desired goals or objectives. Typically, this identification and evaluation process is accomplished with the aid of optimization and simulation models. While optimization methods are designed to provide preferred values of system design and operating policy variables—values that will lead to the highest levels of system performance—they are often used to eliminate the clearly inferior options. Using optimization for a preliminary screening followed by more detailed and accurate simulation is the primary way we have, short of actually building physical models, of estimating effective system designs and operating policies. This chapter introduces and illustrates the art of optimization model development and use in analyzing water resources systems. The models and methods introduced in this chapter are extended in subsequent chapters.

4.1 Introduction

This chapter introduces some optimization modeling approaches for identifying ways of satisfying specified goals or objectives. The modeling 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 optimization methods and approaches. This is not a text on the state of the art of optimization modeling. More realistic and more complex problems usually require much bigger and more complex models than those developed and discussed in this chapter, but these bigger and more complex models are often based on the principles and techniques introduced here.

The emphasis here is on the art of model development—just how one goes about constructing and solving optimization models that will provide information useful for addressing and perhaps even solving particular problems. It is unlikely anyone will ever use any of the specific models developed in this or other chapters simply because the specific examples used to illustrate the approach to model development and solution will not be the ones they face. However, it is quite likely water resource managers and planners will use these modeling approaches and solution methods to analyze a variety of water resource systems. The particular systems modeled and analyzed here, or any others that could have been used, can be the core of more complex models needed to analyze more complex problems in practice.

Water resources planning and management today is dominated by the use of optimization and simulation models. Computer software is

becoming increasingly available for solving various types of optimization and simulation models. However, no software currently exists that will build models of particular water resource systems. What and what not to include and assume in models requires judgment, experience, and knowledge of the particular problems being addressed, the system being modeled and the decision-making environment—including what aspects can be changed and what cannot. Understanding the contents of this and following chapters and performing the suggested exercises at the end of each chapter can only be a first step toward gaining some judgment and experience in model development.

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. Those familiar with this subject that is typically covered in *applied engineering economics* courses can skip this next section.

4.2   Comparing Time Streams of Economic Benefits and Costs

All of us make decisions that involve future benefits and costs. The extent to which we value future benefits or costs compared to present benefits or costs is reflected by what is called a *discount rate*. While economic criteria are only one aspect of everything we consider when making decisions, they are often among the important ones. Economic evaluation methods involving discount rates can be used to consider and compare alternatives characterized by various benefits and costs that are expected to occur now and in the future. This section offers a quick and basic review of the use of discount rates that enable comparisons of alternative time series of benefits and costs. Many economic optimization models incorporate discount rates in their economic objective functions.

Engineering economic methods typically focus on the comparison of a discrete set of mutually exclusive alternatives (only one of which can be selected) each characterized by a time series of benefits and costs. Using various methods involving the discount rate, the time series of benefits and costs are converted to a single net benefit that can be compared with other such net benefits in order to identify the one that is best. The values of the decision variables (e.g., the design and operating policy variable values) are known for each discrete alternative being considered. 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 and the discrete alternatives being considered.

The optimization methods introduced in the following sections of this chapter extend those engineering economics methods. Some methods are discrete, some are continuous. Continuous optimization methods, such as the model defined by Eqs. 3.1–3.3 in Sect. 3.2 of the previous chapter can identify the "best" tank design directly without having to identify and compare numerous discrete, mutually exclusive alternatives. Just how such models can be solved will be discussed later in this chapter. For now, consider the comparison of alternative discrete plans p having different benefits and costs over time.

Let the net benefit generated at the end of time period t by plan p be designated simply as $B^p(t)$. Each plan is characterized by the time stream of net benefits it generates over its planning period T_p.

$$\{B^p(1), B^p(2), B^p(3), \ldots, B^p(T_p) \qquad (4.1)$$

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

4.2.1   Interest Rates

Fundamental to the conversion of a time series of incomes and costs to an equivalent single value, so that it can be compared to other equivalent single values of other time series, 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 owed 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.

Any interest earned by an investor or paid by a debtor depends on the size of the loan, the duration of the loan, and the interest rate. 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 compared to the rate reflecting only the time value of money in a secure and inflation-free environment.

4.2.2   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 each amount in 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 a value 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

$$V_n = V_0(1+r)^n \qquad (4.2)$$

This results from $V_1 = V_0/(1+r)$ at the end of period 1, $V_2 = V_1/(1+r) = V_0(1+r)^2$ at the end of period 2, and so on until at the end of period n.

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

$$V_0 = V_n/(1+r)^n \qquad (4.3)$$

Equation 4.3 is the basic compound interest discounting relation needed to determine the present value at the beginning of period 1 (or end of period 0) 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 $V^p(t)$ 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, i.e., from $t = 1$ to $t = T_p$.

$$V_0^p = \sum_{t=1,T_p} V^p(t)/(1+r)^t \qquad (4.4)$$

The present value of the net benefits achieved by two or more plans having the same economic

planning horizons T_p 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 using present values is possible if all plans have the same planning horizon or if funds remaining at the end of the shorter planning horizon are invested for the remaining time up until the longer planning horizon at the same interest rate r.

4.2.3 Equivalent Annual Value

If the lives of various plans differ, but the same plans will be repeated on into the future, then one need to 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 Eq. 4.4. The equivalent constant annual benefits, V^p, all discounted to the present must equal the present value, V_0^p.

$$V_0^p = \sum_{t=1,T_p} V^p/(1+r)^t \quad \text{or}$$
$$V^p = V_0^p / \sum_{t=1,T_p} 1/(1+r)^t \tag{4.5}$$

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

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

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

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

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.

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

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 analyzing 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. Each of the equations above is applicable to discrete alternatives whose decision variables (investments over time) are known. The equations are used to identify the best alternative from a set of mutually exclusive alternatives whose decision variable values are known. More detailed discussions of the application of
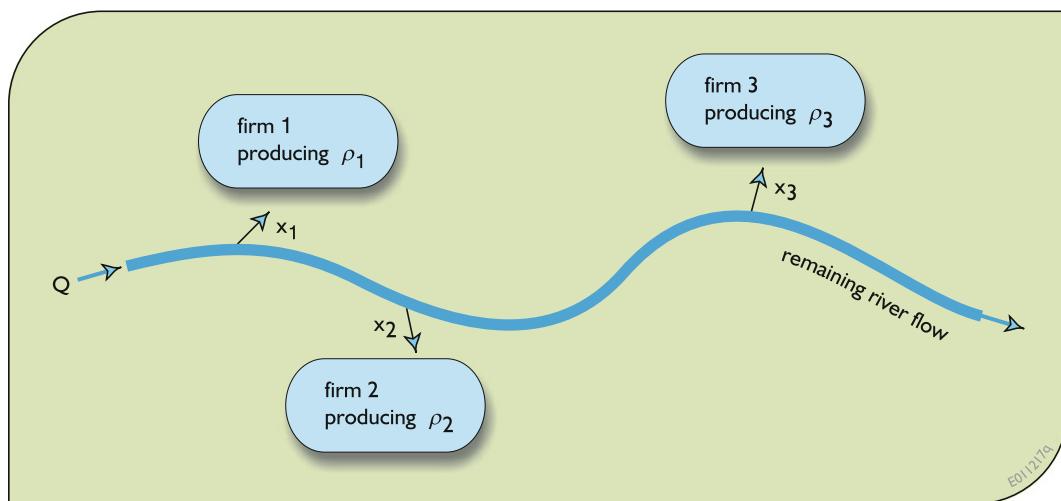
engineering economics are contained in numerous texts on the subject. In the next section, we introduce methods that can identify the best alternative among those whose decision variable values are not known. For example, engineering economic methods can identify, for example, the most cost-effective tank from among those whose dimension values have been previously selected. The optimization methods that follow can identify directly the values of the dimensions of most cost-effective tank.

4.3 Nonlinear Optimization Models and Solution Procedures
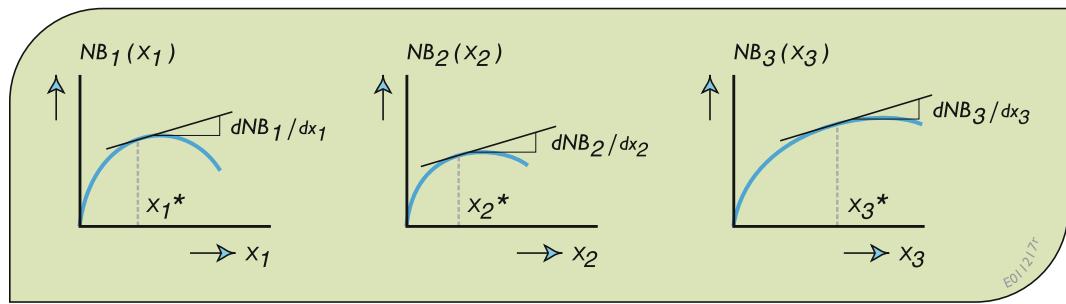
Constrained optimization involves finding the values of decision variables given specified relationships that have to be satisfied. Constrained optimization is also called *mathematical programming*. Mathematical programming techniques include calculus-based Lagrange multipliers and various methods for solving linear and nonlinear models including 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 that in turn is dependent on the system being analyzed. Individuals tend to construct models in a way that will allow them to use a particular optimization technique they think is best. Thus, it pays to be familiar with various types of optimization methods since no one method is best for all optimization problems. Each has its strengths and limitations. The remainder of this chapter introduces and illustrates the application of some of the most common constrained optimization techniques used in water resources planning and management.

Consider a river from which diversions are made to three water-consuming firms that belong to the same corporation, as illustrated in Fig. 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, $\sum_j \mathrm{NB}_j(x_j)$, obtained from all three firms. The total amount of water available is constrained or limited to a quantity of Q.

Fig. 4.1 Three water-using firms obtain water from a river. The amounts x_j allocated to each firm j will depend on the river flow Q

Fig. 4.2 Concave net benefit functions for three water users, j, and their slopes at allocations x_j

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

$$NB_1(x_1) = 6x_1 - x_1^2 \qquad (4.9)$$

$$NB_2(x_2) = 7x_2 - 1.5x_2^2 \qquad (4.10)$$

$$NB_3(x_3) = 8x_3 - 0.5x_3^2 \qquad (4.11)$$

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

4.3.1   Solution Using Calculus

Calculus can be used to find the allocations that maximize each user's net benefits, simply by finding where the slope or derivative of the net benefit function for each firm equals zero. The derivative, $dNB(x_1)/dx_1$, of the net benefit function for Firm 1 is $(6 - 2x_1)$ and hence the allocation to Firm 1 that maximizes its net benefits would be 6/2 or 3. The corresponding 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 8 units of flow are available for all three firms and 2
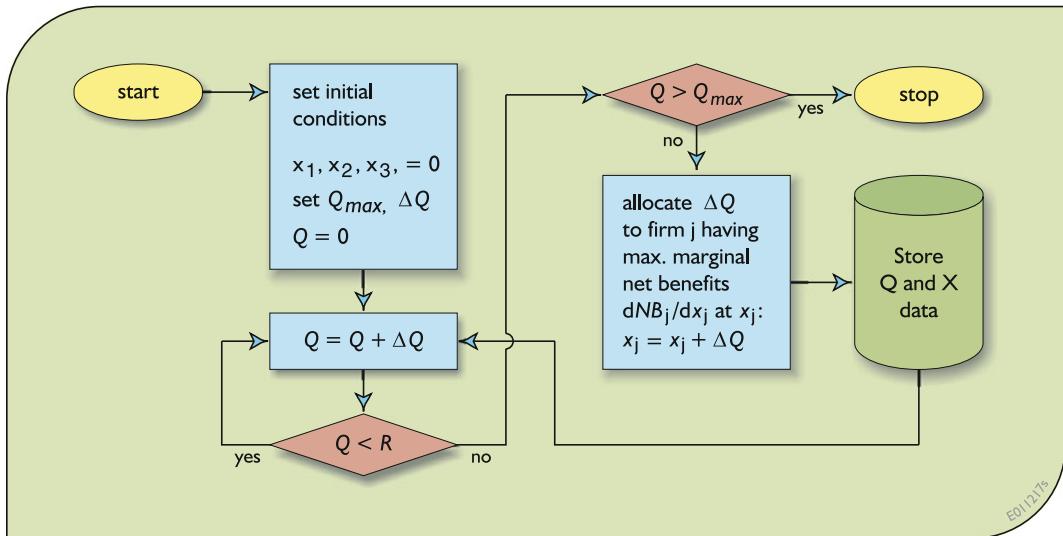
units must remain in the river. Introducing this constraint renders the previous solution infeasible. In this case we want to find the allocations that maximize the total net benefits obtained from all firms subject to having only 6 flow units available for allocations. Using simple calculus will not suffice.

4.3.2   Solution Using Hill Climbing

One approach for finding, at least approximately, 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 successive increment so as to get the maximum additional net benefit from that incremental amount of water. This procedure works in this example because each of the net benefit functions is concave; in other words, the marginal benefits decrease as the allocation increases. This procedure is illustrated by the flow diagram in Fig. 4.3.

Table 4.1 lists the results of applying the procedure shown in Fig. 4.3 to the problem when (a) only 8 and (b) only 20 flow units are available. Here a minimum river flow of 2 is required and is to be satisfied, when possible, before any allocations are made to the firms.

The hill-climbing method illustrated in Fig. 4.3 and Table 4.1 assigns each incremental

Fig. 4.3 Steepest hill-climbing approach for finding allocation of a flow Q_{max} to the three firms, while meeting minimum river flow requirements R

flow ΔQ to the use that yields the largest additional (marginal) net benefit. An allocation is optimal for any total flow Q when the marginal net benefits from each nonzero allocation are equal, or as close to each other as possible given the size of the increment ΔQ. In this example, with a ΔQ of 1 and Q_{max} of 8, it just happens that the marginal net benefits associated with each allocation are all equal (to 4). The smaller the ΔQ, the more precise will be the optimal allocations in each iteration, as shown in the lower portion of Table 4.1, where ΔQ 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 for any particular value of Q is shown in Fig. 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.

4.3.3 Solution Using Lagrange Multipliers

4.3.3.1 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

Table 4.1 Hill-climbing iterations for finding allocations that maximize total net benefit given a flow of Q_{max} and a required (minimum) streamflow of $R = 2$
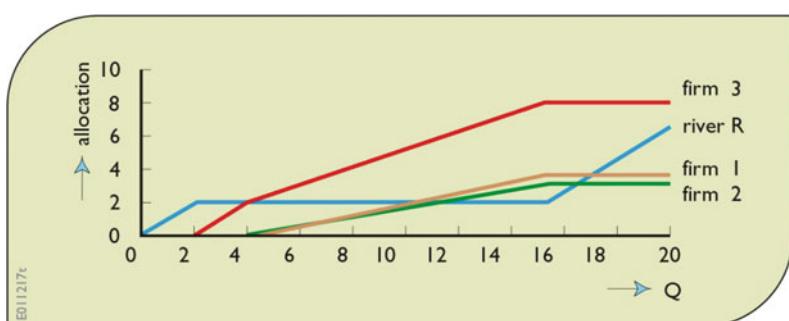
$Q_{max} = 8$; $Q_i = 0$; $\Delta Q = 1$;   river flow $R \geq \min \{Q, 2\}$

iteration i	Q_i	allocations. R, x_j				marginal net benefits			new allocations $Q_i + \Delta Q$					total net benefits
		R	x_1	x_2	x_3	$6-2x_1$	$7-3x_2$	$8-x_3$	R	x_1	x_2	x_3		$\Sigma_j\, NB_j(x_1)$
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	–	–	–	–	–	---

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

Q	allocations. R, x_j				marginal net benefits			total net benefits
	R	x_1	x_2	x_3	$6-2x_1$	$7-3x_2$	$8-x_3$	$\Sigma_j\, NB_j(x_1)$
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

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

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

$$P_1(x_1) = 0.4(x_1)^{0.9} \qquad (4.12)$$

$$P_2(x_2) = 0.5(x_2)^{0.8} \qquad (4.13)$$

$$P_3(x_3) = 0.6(x_3)^{0.7} \qquad (4.14)$$

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

$$C_1 = 3(P_1(x_1))^{1.3} \qquad (4.15)$$

$$C_2 = 5(P_2(x_2))^{1.2} \qquad (4.16)$$

$$C_3 = 6(P_3(x_3))^{1.15} \qquad (4.17)$$

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 unit price and the amount that will be demanded and sold. These are the demand functions for that product. These unit price or demand functions are shown in Fig. 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.
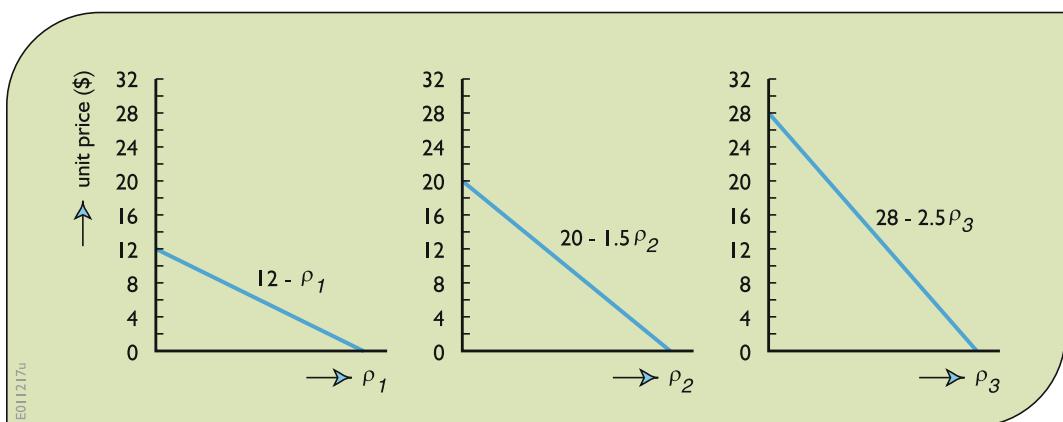
The optimization problem is to find the water allocations, the production levels, and the unit prices that together maximize the total net benefit obtained from all three firms. The water allocations plus the amount that must remain in the river, R, cannot exceed the total amount of water Q available.

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 Fig. 4.4, showing the three best allocations given any value of Q. This of course assumes the firms can adjust to varying allocations. In reality this may not be the case (Chapter 9 examines this problem using more realistic benefit functions that reflect the degree to which firms can adapt to changing inputs over time.)

The model:

$$\text{Maximize Net\_benefit} \qquad (4.18)$$

Fig. 4.5 Unit prices that will guarantee the sale of the specified amounts of products p_j produced in each of the three firms (linear functions are assumed in this example for simplicity)

Subject to
Definitional constraints:

Net_benefit = Total_return − Total_cost

$$(4.19)$$

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

$$(4.20)$$

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

$$(4.21)$$

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

$$p_1 = 0.4(x_1)^{0.9} \qquad (4.22)$$

$$p_2 = 0.5(x_2)^{0.8} \qquad (4.23)$$

$$p_3 = 0.6(x_3)^{0.7} \qquad (4.24)$$

Water-allocation restriction

$$R + x_1 + x_2 + x_3 = Q \qquad (4.25)$$

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:

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

$$(4.26)$$

Derivatives:

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

$$(4.27)$$

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

$$(4.28)$$

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

$$(4.29)$$

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 product production at, one or more of the three firms.

If the total available amount of water is less than that desired, constraint Eq. 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 Eqs. 4.22–4.25 need to be added. These can be rewritten as equalities since they will be binding. Equation 4.25 in this case can always be an equality since any excess water will be allocated to the river, R.

To consider values of Q that are less than the desired 38.3 units, constraints 4.22–4.25 can be included in the objective function, Eq. 4.26, once the right-hand side has been subtracted from the left-hand side so that they equal 0. We set this function equal to L.

$$L = [(12 - p_1)p_1 + (20 - 1.5p_2)p_2 + (28 - 2.5p3)p_3]$$
$$- \left[3(p_1)^{1.30} + 5(p_2)^{1.20} + 6(p_3)^{1.15}\right]$$
$$- \lambda_1\left[p_1 - 0.4(x_1)^{0.9}\right] - \lambda_2\left[p_2 - 0.5(x_2)^{0.8}\right]$$
$$- \lambda_3\left[p_3 - 0.6(x_3)^{0.7}\right] - \lambda_4[R + x_1 + x_2 + x_3 - Q]$$

$$(4.30)$$

Since each of the four constraint Eqs. 4.22–4.25 included in Eq. 4.30 equals zero, each can be multiplied by a variable λ_i without changing the value of Eq. 4.30, or equivalently, Eq. 4.26. These unknown variables λ_i are called the Lagrange multipliers of constraints i. The value of each multiplier, λ_i, is the marginal value of the original objective function, Eq. 4.26, with respect to a change in the value of the amount produced, p, or in the case of constraint Eq. 4.25, the amount of water available, Q. We will show this shortly.

Differentiating Eq. 4.30 with respect to each of the ten unknowns and setting the resulting equations to 0 yields:

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

$$(4.31)$$

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

$$(4.32)$$

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

$$(4.33)$$

$$\partial L/\partial x_1 = 0 = \lambda_1 0.9(0.4)x_1^{-0.1} - \lambda_4 \quad (4.34)$$

$$\partial L/\partial x_2 = 0 = \lambda_2 0.8(0.5)x_2^{-0.2} - \lambda_4 \quad (4.35)$$

$$\partial L/\partial x_3 = 0 = \lambda_3 0.7(0.6)x_3^{-0.3} - \lambda_4 \quad (4.36)$$

$$\partial L/\partial \lambda_1 = 0 = p_1 - 0.4(x_1)^{0.9} \quad (4.37)$$

$$\partial L/\partial \lambda_2 = 0 = p_2 - 0.5(x_2)^{0.8} \quad (4.38)$$

$$\partial L/\partial \lambda_3 = 0 = p_3 - 0.6(x_3)^{0.7} \quad (4.39)$$

$$\partial L/\partial \lambda_4 = 0 = R + x_1 + x_2 + x_3 - Q \quad (4.40)$$

These ten equations are the conditions necessary for a solution that maximizes Eq. 4.30, or equivalently 4.26. 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 using LINGO) are shown in Table 4.2. (A free demo version of LINGO can be obtained (downloaded) from www.LINDO.com.)

4.3.3.2 Meaning of Lagrange Multiplier λ

In this example, Eq. 4.30 is the objective function that is to be maximized. It is maximized or

Table 4.2 Solutions to Eqs. 4.31–4.40

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	λ_1	λ_2	λ_3	λ_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

minimized by equating to zero each of its partial derivatives with respect to each unknown variable. Equation 4.30 consists of the original net benefit function plus each constraint i multiplied by a weight or multiplier λ_i. This equation is expressed in monetary units. 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 λ_i associated with these constraints are expressed in monetary units per constraint units. In this example, the multipliers λ_1, λ_2, and λ_3 represent the change in the total net benefit value of the objective function (Eq. 4.26) per unit change in the products p_1, p_2, and p_3 produced. The multiplier λ_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, λ_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, is a concave function of those quantities, as illustrated in Fig. 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.

$$\text{Maximize (or minimize)}\, F(X) \quad (4.41)$$

subject to constraints

$$g_i(X) = b_i \quad i = 1, 2, 3, \ldots, m, \quad (4.42)$$

where X is the vector of all x_j. The Lagrange function $L(X, \lambda)$ is formed by combining Eq. 4.42, each equaling zero, with the objective function of Eq. 4.41.

$$L(X, \lambda) = F(X) - \sum_i \lambda_i(g_i(X) - b_i) \quad (4.43)$$

Solutions of the equations $\partial L/\partial x_j = 0$ for all decision variables x_j and $\partial L/\partial \lambda_i = 0$ for all constraints g_i 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 λ_i at the optimum solution can be derived by manipulation of $\partial L/\partial \lambda_i = 0$. Taking the partial derivative of the Lagrange function, Eq. 4.43, with respect to an unknown variable x_j and setting it to zero results in

$$\partial L/\partial x_j = 0 = \partial F/\partial x_j - \sum_i \lambda_i \partial(g_i(X))/\partial x_j \quad (4.44)$$

Multiplying each term by ∂x_j yields

$$\partial F = \sum_i \lambda_i \partial(g_i(X)) \quad (4.45)$$

Dividing each term by ∂b_k associated with a particular constraint, say k, defines the meaning of λ_k.

$$\partial F/\partial b_k = \sum_i \lambda_i \partial(g_i(X))/\partial b_k = \lambda_k \quad (4.46)$$

Equation 4.46 follows from the fact that $\partial(g_i(\mathbf{X}))/\partial b_k$ equals 0 for constraints $i \neq k$ and equals 1 for the constraint $i = k$. The latter is true since $b_i = g_i(X)$ and thus $\partial(g_i(X)) = \partial b_i$.

From Eq. 4.46, each multiplier λ_i is the marginal change in the original objective function $F(X)$ with respect to a change in the constant b_i associated with the constraint i. For nonlinear problems, it is the slope of the objective function plotted against the value of b_i.

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 $g_i(X) \leq b_i$ a squared slack variable S_i^2 can be added to the left-hand side to make it an equation $g_i(X) + S_i^2 = b_i$. For a

greater-than-or-equal constraint $g_i(X) \geq b_i$ a squared surplus variable S_i^2 can be subtracted from the left-hand side to make it an equation $g_i(X) - S_i^2 = b_i$. These slack or surplus variables are squared to ensure they are nonnegative, and also to make them appear in the differential equations.

$$\partial L/\partial S_i = 0 = -2S_i\lambda_i = S_i\lambda_i \qquad (4.47)$$

Equation 4.47 shows that either the slack or surplus variable, S, or the multiplier, λ, will always be zero. If the value of the slack or surplus variable S is nonzero, the constraint is redundant. The optimal solution will not be affected by the constraint. Small changes in the values, b, of redundant constraints will not change the optimal value of the objective function $F(X)$. Conversely, if the constraint is binding, the value of the slack or surplus variable S will be zero. The multiplier λ can be nonzero if the value of the function $F(X)$ is sensitive to the constraint value b.

The solution of the set of partial differential Equations Eqs. 4.47 often involves a trial-and-error process, equating to zero a λ or a S for each inequality constraint and solving the remaining equations, if possible. This tedious procedure, along with the need to check boundary solutions when nonnegativity conditions are imposed, detracts from the utility of classical Lagrange multiplier methods for solving all but relatively simple water resources planning problems.

4.4 Dynamic Programming

The water-allocation problems in the previous section assumed a net-benefit function for each water-using firm. In those examples, these functions were continuous and differentiable, 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 modeling is used to refine the solutions identified using dynamic programming.
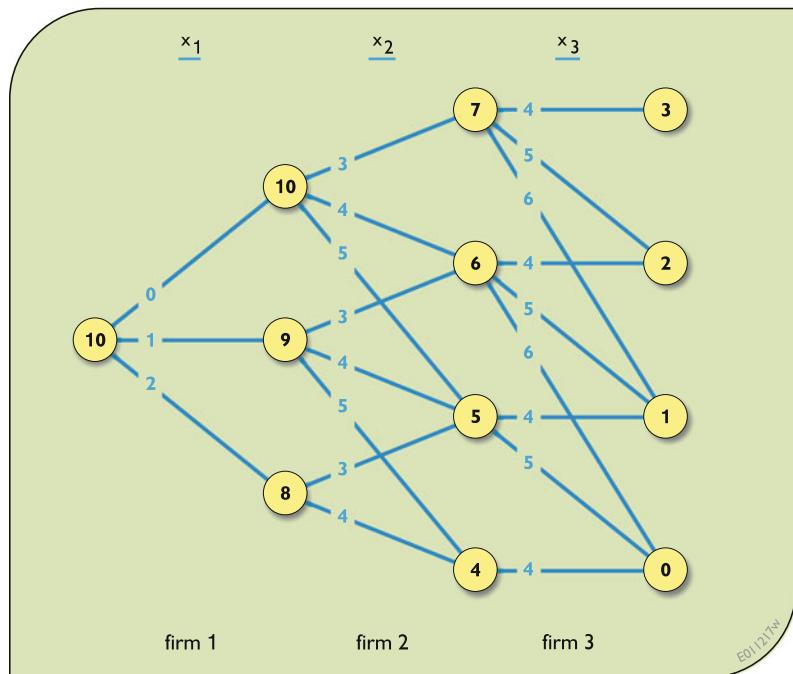
Dynamic programming is an approach that divides the original optimization problem, with all of its variables, into a set of smaller optimization problems, each of which needs to be solved before the overall optimum solution to the original problem can be identified. The water supply allocation problem, 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.

4.4.1 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, or go from, any node in that network. The nodes represent possible discrete states of the system that can exist and the links represent the decisions one could make to get from one state (node) to another. Figure 4.6 illustrates a portion of such a network for the three-firm allocation problem shown in Fig. 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 Fig. 4.6. The states of the system are the nodes and the values of the states are the numbers in the

Fig. 4.6 A network representing some of the possible integer allocations of water to three water-consuming firms *j* assuming 10 units of water are available. The *circles* or *nodes* represent the discrete quantities of water available to users not yet allocated any water, and the links represent feasible allocation decisions x_j to the next firm *j*

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 different components (e.g., firms) or time periods. Links between (or connecting) initial and final states represent decisions. The links in this example represent possible allocation decisions for each of the three different firms. Each stage is a separate firm. Each state is an amount of water that remains to be allocated in the remaining stages.
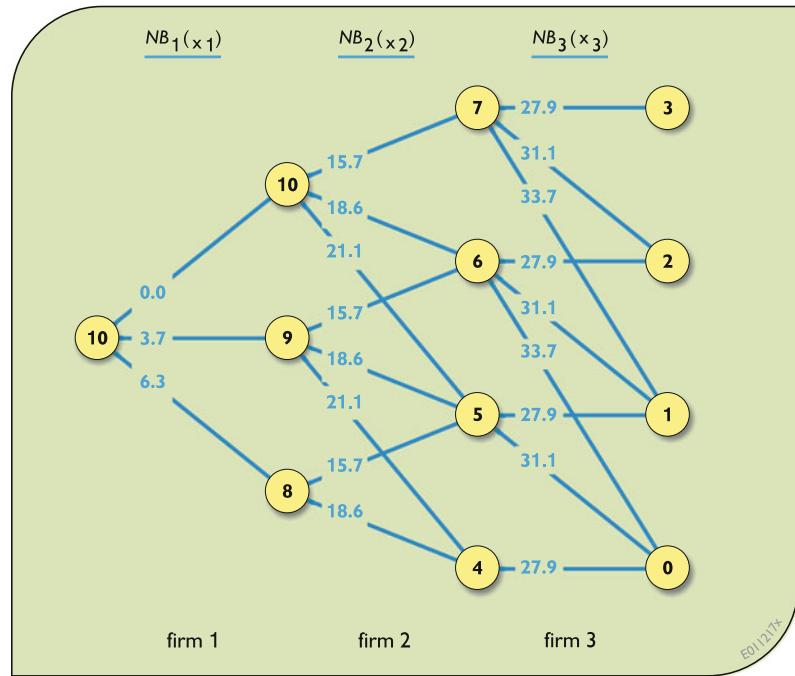
Each link connects two nodes, the left node value indicating the state of a system before a decision is made, and the right node value indicating the state of a system after a decision is made. In this case, the state of the system is the amount of water available to allocate to the remaining firms.

In the example shown in Fig. 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 would not know this solution before solving for it using dynamic programming, but since we do we can reduce the complexity (number of nodes and links) 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 integer 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 (noninteger) discretizations would involve even more nodes and connecting links.

The links of Fig. 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 of the link. The number in the right node is

Fig. 4.7 Network as in Fig. 4.6 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 integer allocation decisions

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 S_1 that can be allocated to all three firms, after allocating x_1 to Firm 1 what remains is S_2:

$$S_1 - x_1 = S_2 \qquad (4.48)$$

Allocating x_2 to Firm 2, leaves S_3.

$$S_2 - x_2 = S_3 \qquad (4.49)$$

Finally, allocating x_3 to Firm 3 leaves S_4.

$$S_3 - x_3 = S_4 \qquad (4.50)$$

Figure 4.6 shows the different values of each of these states, S_j, and decision variables x_j beginning with a quantity $S_1 = Q - R = 10$. Our task is to find the best path through the network, beginning at the leftmost 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 shows the same network as in Fig. 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

$$NB_1(x_1) = \text{maximum}(12 - p_1)p_1 - 3(p_1)^{1.30}$$
$$\text{where } p_1 \le 0.4(x_1)^{0.9}$$

$$(4.51)$$

$$NB_2(x_2) = \text{maximum}(20 - 1.5p_2)p_2 - 5(p_2)^{1.20}$$
$$\text{where } p_2 \le 0.5(x_2)^{0.8}$$

$$(4.52)$$

$$NB_3(x_3) = \text{maximum}(28 - 2.5p_3)p_3 - 6(p_3)^{1.15}$$
$$\text{where } p_3 \le 0.6(x_3)^{0.7}$$

$$(4.53)$$

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 vertical 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

$$\text{Maximize} \quad \sum_j \text{NB}_j(x_j), \qquad (4.54)$$

where $\text{NB}_j(x_j)$ is the net benefit associated with an allocation of x_j to firm j. Equations 4.51–4.53 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 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 can be viewed as a multistage 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., for each node representing an 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 $j, j + 1, j + 2 \ldots$? 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 leftmost 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. We will revisit this issue when we get to reservoir operation where the stages are time periods.

4.4.2   Backward-Moving Solution Procedure

Consider the network in Fig. 4.7. Again, the nodes represent the discrete states—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 Fig. 4.7 include:
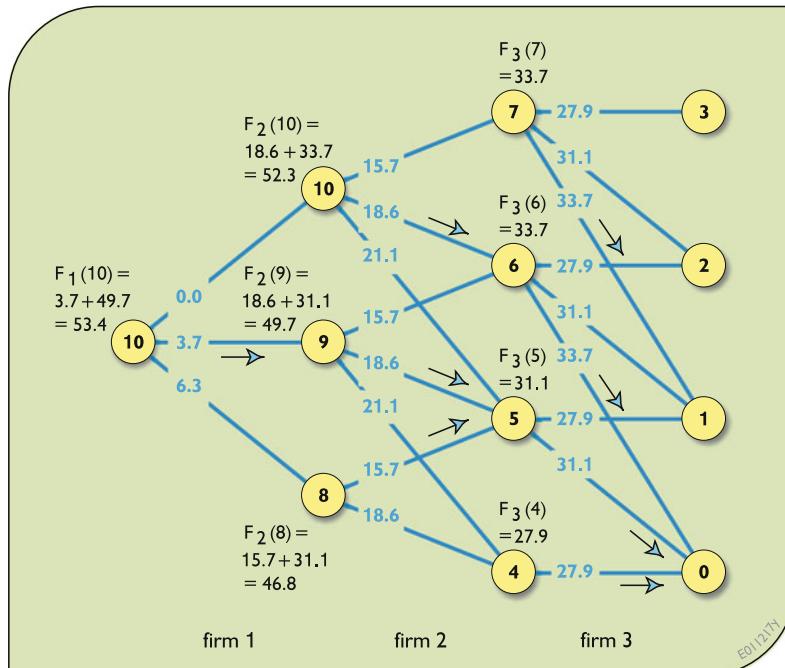
$$F_3(7) = \text{Maximum}\{\text{NB}_3(x_3)\}$$
$$x_3 \leq 7, \text{ the total flow available.}$$
$$4 \leq x_3 \leq 6, \text{ the allowable range of allocations}$$
$$= \text{Maximum}\{27.9, 31.1, 33.7\} = 33.7 \text{ when } x_3 = 6 \tag{4.55}$$

$$F_3(6) = \text{Maximum}\{\text{NB}_3(x_3)\}$$
$$x_3 \leq 6$$
$$4 \leq x_3 \leq 6$$
$$= \text{Maximum}\{27.9, 31.1, 33.7\} = 33.7 \text{ when } x_3 = 6 \tag{4.56}$$

$$F_3(5) = \text{Maximum}\{\text{NB}_3(x_3)\}$$
$$x_3 \leq 5$$
$$4 \leq x_3 \leq 6$$
$$= \text{Maximum}\{27.9, 31.1\} = 31.1 \text{ when } x_3 = 5 \tag{4.57}$$

$$F_3(4) = \text{Maximum}\{\text{NB}_3(x_3)\}$$
$$x_3 \leq 4$$
$$4 \leq x_3 \leq 6$$
$$= \text{Maximum}\{27.9\} = 27.9 \text{ when } x_3 = 4 \tag{4.58}$$

These computations are shown on the network in Fig. 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 Fig. 4.8 the links representing the decisions or allocations that result in the largest net benefits are shown with arrows.

Fig. 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 allocation policy that maximizes 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

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 Eq. 4.49, 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 Fig. 4.8:

$$F_2(10) = \text{Maximum}\{\text{NB}_2(x_2) + F_3(S_3 = 10 - x_2)\}$$
$$x_2 \leq 10$$
$$3 \leq x_2 \leq 5$$
$$= \text{Maximum}\{15.7 + 33.7, 18.6$$
$$+ 33.7, 21.1 + 31.1\} = 52.3 \text{ when } x_2 = 4$$
$$(4.59)$$

$$F_2(9) = \text{Maximum}\{\text{NB}_2(x_2) + F_3(S_3 = 9 - x_2)\}$$
$$x_2 \leq 9$$
$$3 \leq x_2 \leq 5$$
$$= \text{Maximum}\{15.7 + 33.7, 18.6$$
$$+ 31.1, 21.1 + 27.9\} = 49.7$$
$$\text{when } x_2 = 4$$
$$(4.60)$$

$$F_2(8) = \text{Maximum}\{\text{NB}_2(x_2) + F_3(S_3 = 8 - x_2)\}$$
$$x_2 \leq 8$$
$$3 \leq x_2 \leq 5 \text{ (assume 4 instead of 5 since both}$$
$$\text{will not affect optimal solution)}$$
$$= \text{Maximum}\{15.7 + 31.1, 18.6$$
$$+ 27.9\} = 46.8 \text{ when } x_2 = 3$$
$$(4.61)$$

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 beginning stage 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 given 10 units of water available, is

$$F_1(10) = \text{Maximum}\{\text{NB}_1(x_1) + F_2(S_2 = 10 - x_1)\}$$
$$x_1 \leq 10$$
$$0 \leq x_1 \leq 2$$
$$= \text{Maximum}\{0 + 52.3, 3.7$$
$$+ 49.7, 6.3 + 46.8\} = 53.4 \text{ when } x_1 = 1$$
$$(4.62)$$

The value of $F_1(10)$ in Eq. 4.62 is the same as the value of Eq. 4.54. This value is the maximum net benefits obtainable from allocating the available 10 units of water. From Eq. 4.62 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 Eq. 4.60 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, Eq. 4.57 tells us we should allocate all 5 units to Firm 3. All this is illustrated in Fig. 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 allocation values is 1, 4, and 5, respectively.

To summarize, a dynamic programming model was developed for the following problem:

$$\text{Maximize Net\_benefit} \qquad (4.63)$$

Subject to

$$\text{Net\_benefit} = \text{Total\_return} - \text{Total\_cost} \tag{4.64}$$

$$\text{Total\_return} = (12 - p_1)p_1 + (20 - 1.5p_2)p_2 \\ + (28 - 2.5p_3)p_3 \tag{4.65}$$

$$\text{Total\_cost} = 3(p_1)^{1.30} + 5(p_2)^{1.20} + 6(p_3)^{1.15} \tag{4.66}$$

$$p_1 \leq 0.4(x_1)^{0.9} \tag{4.67}$$

$$p_2 \leq 0.5(x_2)^{0.8} \tag{4.68}$$

$$p_3 \leq 0.6(x_3)^{0.7} \tag{4.69}$$

$$x_1 + x_2 + x_3 \leq 10 \tag{4.70}$$

The discrete dynamic programming version of this problem required discrete states S_j representing the amount of water available to allocate to firms j, $j + 1$, …. 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 backward as indicated in Eqs. 4.71–4.73.

$$F_3(S_3) = \text{maximum}\{NB_3(x_3)\} \text{ over all } x_3 \leq S_3, \\ \text{for all discrete } S_3 \text{ values between } 0 \text{ and } 10 \tag{4.71}$$

$$F_2(S_2) = \text{maximum}\{NB_2(x_2) + F_3(S_3)\} \\ \text{over all } x_2 \leq S_2 \text{ and } S_3 = S_2 - x_2, \quad 0 \leq S_2 \leq 10 \tag{4.72}$$

$$F_1(S_1) = \text{maximum}\{NB_1(x_1) + F_2(S_2)\} \\ \text{over all } x_1 \leq S_1 \text{ and } S_2 = S_1 - x_1 \text{ and } S_1 = 10 \tag{4.73}$$

The values of each $NB_j(x_j)$ are obtained from Eqs. 4.51 to 4.53.

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 4.71 need to be solved before Eqs. 4.72 can be solved, and Eqs. 4.72 need to be solved before Eqs. 4.73 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 nonlinear optimization model defined by Eqs. 4.63–4.70 and the dynamic programming model defined by the recursive Eqs. 4.71–4.73. Note that $F_3(S_3)$ in Eq. 4.71 is the same as

$$F_3(S_3) = \text{Maximum } NB_3(x_3) \tag{4.74}$$

Subject to

$$x_3 \leq S_3, \tag{4.75}$$

where $NB_3(x_3)$ is defined in Eq. 4.53.

Similarly, $F_2(S_2)$ in Eq. 4.72 is the same as

$$F_2(S_2) = \text{Maximum } NB_2(x_2) + NB_3(x_3) \tag{4.76}$$

Subject to

$$x_2 + x_3 \leq S_2, \tag{4.77}$$

where $NB_2(x_2)$ and $NB_3(x_3)$ are defined in Eqs. 4.52 and 4.53.

Finally, $F_1(S_1)$ in Eq. 4.73 is the same as

$$F_1(S_1) = \text{Maximum } NB_1(x_1) + NB_2(x_2) + NB_3(x_3) \tag{4.78}$$

Subject to

$$x_1 + x_2 + x_3 \leq S_1 = 10, \tag{4.79}$$

where $NB_1(x_1)$, $NB_2(x_2)$, and $NB_3(x_3)$ are defined in Eqs. 4.51–4.53.

Alternatively, $F_3(S_3)$ in Eq. 4.71 is the same as

$$F_3(S_3) = \text{Maximum}(28 - 2.5p_3)p_3 - 6(p_3)^{1.15}$$
(4.80)

Subject to

$$p_3 \le 0.6(x_3)^{0.7}$$
(4.81)

$$x_3 \le S_3$$
(4.82)

Similarly, $F_2(S_2)$ in Eq. 4.72 is the same as

$$F_2(S_2) = \text{Maximum}(20 - 1.5p_2)p_2$$
$$+ (28 - 2.5p_3)p3 - 5(p2)^{1.20} - 6(p3)^{1.15}$$
(4.83)

Subject to

$$p_2 \le 0.5(x_2)^{0.8}$$
(4.84)

$$p_3 \le 0.6(x_3)^{0.7}$$
(4.85)

$$x_2 + x_3 \le S_2$$
(4.86)

Finally, $F_1(S_1)$ in Eq. 4.73 is the same as

$$F_1(S_1) = \text{Maximum}(12 - p_1)p_1$$
$$+ (20 - 1.5p_2)p_2 + (28 - 2.5p_3)p_3$$
$$- \left[3(p_1)^{1.30} + 5(p_2)^{1.20} + 6(p_3)^{1.15}\right]$$
(4.87)

Subject to

$$p_1 \le 0.4(x_1)^{0.9}$$
(4.88)

$$p_2 \le 0.5(x_2)^{0.8}$$
(4.89)

$$p_3 \le 0.6(x_3)^{0.7}$$
(4.90)

$$x_1 + x_2 + x_3 \le S_1 = 10$$
(4.91)

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 Eqs. 4.48–4.50, or, in general terms for all firms j, by

$$S_{j+1} = S_j - x_j$$
(4.92)

4.4.3 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 the remaining water, state S_{j+1}. 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 Fig. 4.7 or Fig. 4.8 the allocation decisions, x_1, given a value for S_2 are obvious.

$$f_1(S_2) = \text{maximum}\{NB_1(x_1)\}$$
$$x_1 = 10 - S_2$$
(4.93)

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 x_2 to have made given each final state S_3.

$$f_2(S_3) = \text{maximum}\{NB_2(x_2) + f_1(S_2)\}$$
$$0 \le x_2 \le 10 - S_3 \qquad (4.94)$$
$$S_2 = S_3 + x_2$$

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 x_3 to have made given each final state S_4.

$$f_3(S_4) = \text{maximum}\{NB_3(x_3) + f_2(S_3)\}$$
$$\text{for all discrete } S_4 \text{ between 0 and 10.}$$
$$0 \le x_3 \le 10 - S_4$$
$$S_3 = S_4 + x_3$$
$$(4.95)$$

Figure 4.9 illustrates a portion of the network represented by Eqs. 4.93–4.95, and the $f_j(S_{j+1})$ values.
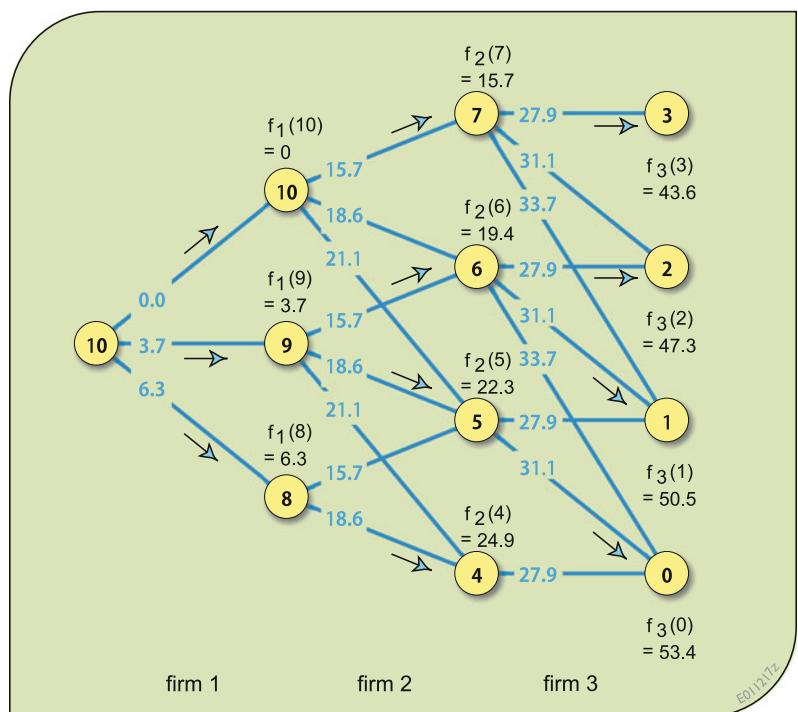
From Fig. 4.9, note the highest total net benefits are obtained by ending with 0 remaining

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. 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.

4.4.4 Numerical Solutions

The application of discrete dynamic programming to most practical problems will usually 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 unless they

Fig. 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

do it by hand. Most computer programs written for solving specific dynamic programming problems create and store the solutions of the recursive equations (e.g., Eqs. 4.93–4.95) 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 Figs. 4.8 and 4.9. The backward solution procedure is used.

Table 4.3 contains the solutions of Eqs. 4.55–4.58 for the third stage. Table 4.4 contains the solutions of Eqs. 4.59–4.61 for the second stage. Table 4.5 contains the solution of Eq. 4.62 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 Fig. 4.5 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,

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 $NB_2(S_2) + F_3(S_3 = S_2 - x_2)$				
	decisions: x_2				
state S_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 $NB_1(S_1) + F_2(S_2 = S_1 - x_1)$				
	decisions: x_2				
state S_1	0	1	2	$F_1(S_1)$	x_1
10	0 + 52.3	3.7 + 49.7	6.3 + 46.8	53.4	1

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.

4.4.5 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 multistate-variable problems.

4.4.6 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 needed in order to begin the process of solving each successive recursive equation.

4.4.7 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

(Fig. 4.1) illustrates the first type of application. The other two applications are presented below.

4.4.7.1   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.

$$\text{Minimize} \quad \sum_t C_t(s_t, x_t), \qquad (4.96)$$

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.

$$s_{t+1} = s_1 + \sum_{\tau=1,t} x_\tau \quad \text{for } t = 1, 2, \ldots, T$$
$$(4.97)$$

Alternatively these equations may be expressed by a series of continuity relationships:

$$s_{t+1} = s_t + x_t \quad \text{for } t = 1, 2, \ldots, T \qquad (4.98)$$

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.
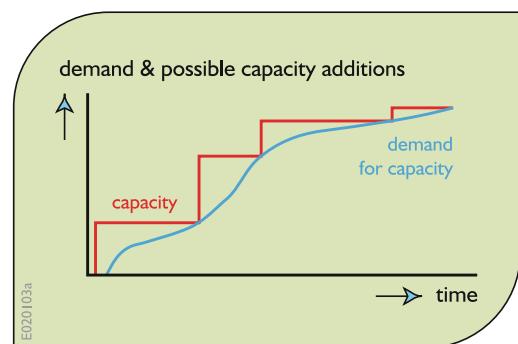
$$s_{t+1} \geq D_t \quad \text{for } t = 1, 2, \ldots, T \qquad (4.99)$$

There may also be constraints on the possible expansions in each period defined by a set Ω_t of feasible capacity additions in each period t:
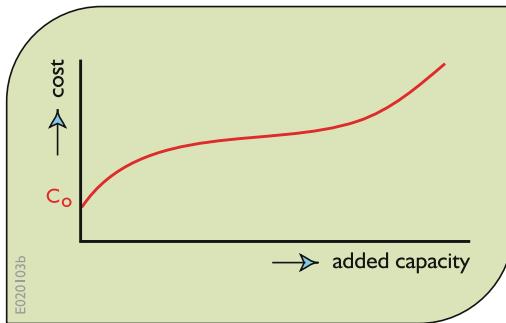
$$x_t \in \Omega_t \qquad (4.100)$$

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 fixed costs and economies of scale, as illustrated in Fig. 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 in the future to meet an increased demand for additional capacity.
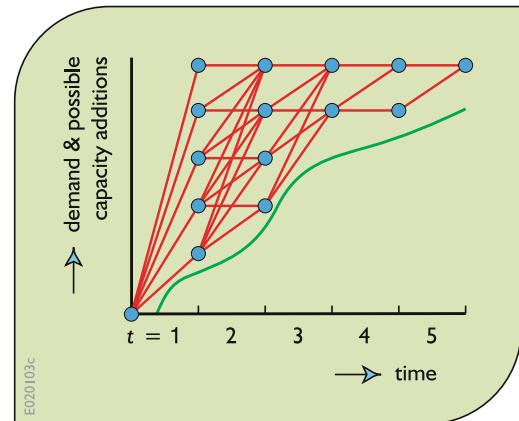
The problem is also important because any estimates made today of future demands, costs and interest rates are likely to be wrong. The future is uncertain. Its uncertainties increase the further the future. Capacity expansion planners need to consider the future if their plans are to be cost-effective and not myopic from assuming

Fig. 4.10 A demand projection (*solid blue line*) and a possible capacity expansion schedule (*red line*) for meeting that projected demand over time

Fig. 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

Fig. 4.12 Network of discrete capacity expansion decisions (*links*) that meet the projected demand

there is no future. Just how far into the future do they need to look? And what about the uncertainty in all future costs, demands, and interest rate 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 Eqs. 4.96–4.100 can be restructured as a multistage 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 Fig. 4.9, as shown in Fig. 4.12. The solid blue circles in Fig. 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

$$f_1(s_2) = \min\{C_1(s_1, x_1)\} \text{ in which the discrete } x_1 = s_2 \text{ and } s_1 = 0 \tag{4.101}$$

Moving to stage $t = 2$, for the final discrete states s_3 ranging from D_2 to D_T,

$$f_2(s_3) = \min\{C_2(s_2, x_2) + f_1(s_2)\}$$
$$\text{over all discrete } x_2 \text{ between } 0 \tag{4.102}$$
$$\text{and } s_3 - D_1 \text{ and } s_2 = s_3 - x_2$$

Moving to stage $t = 3$, for the final discrete states s_4 ranging from D_3 to D_T,

$$f_3(s_4) = \min\{C_3(s_3, x_3) + f_2(s_3)\}$$
$$\text{over all discrete } x_3 \text{ between } 0$$
$$\text{and } s_4 - D_2 \text{ and } s_3 = s_4 - x_3 \tag{4.103}$$

In general for all stages t between the first and last:

$$f_t(s_{t+1}) = \min\{C_t(s_t, x_t) + f_{t-1}(s_t)\}$$
$$\text{over all discrete } x_t \text{ between } 0$$
$$\text{and } s_{t+1} - D_{t-1} \text{ and } s_t = s_{t+1} - x_t \tag{4.104}$$

For the last stage $t = T$ and for the final discrete state $s_{T+1} = D_T$,

$$f_T(s_{T+1}) = \min\{C_T(s_T, x_T) + f_{T-1}(s_T)\}$$
$$\text{over all discrete } x_T$$
$$\text{between } 0 \text{ and } D_T - D_{T-1}$$
$$\text{where } s_T = s_{T+1} - x_T$$

$$(4.105)$$

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.
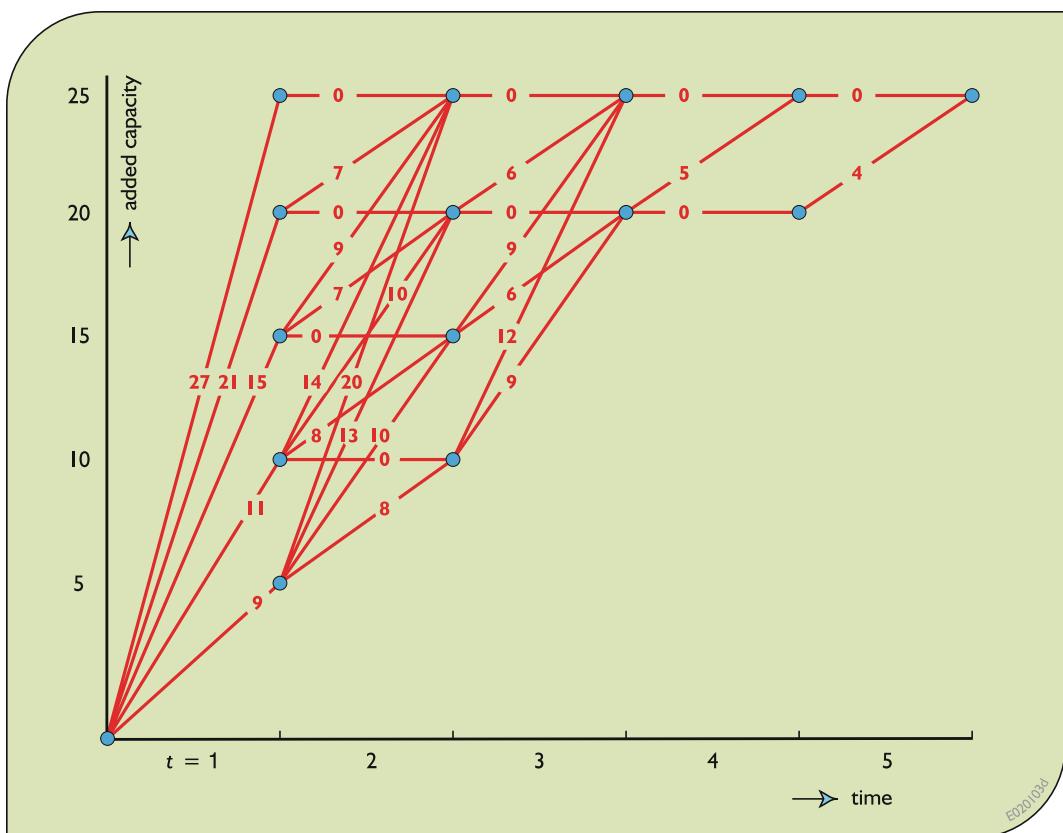
A numerical example

Consider the five-period capacity expansion problem shown in Fig. 4.12. Figure 4.13 is the

same network 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 Fig. 4.14 using the forward-moving algorithm.
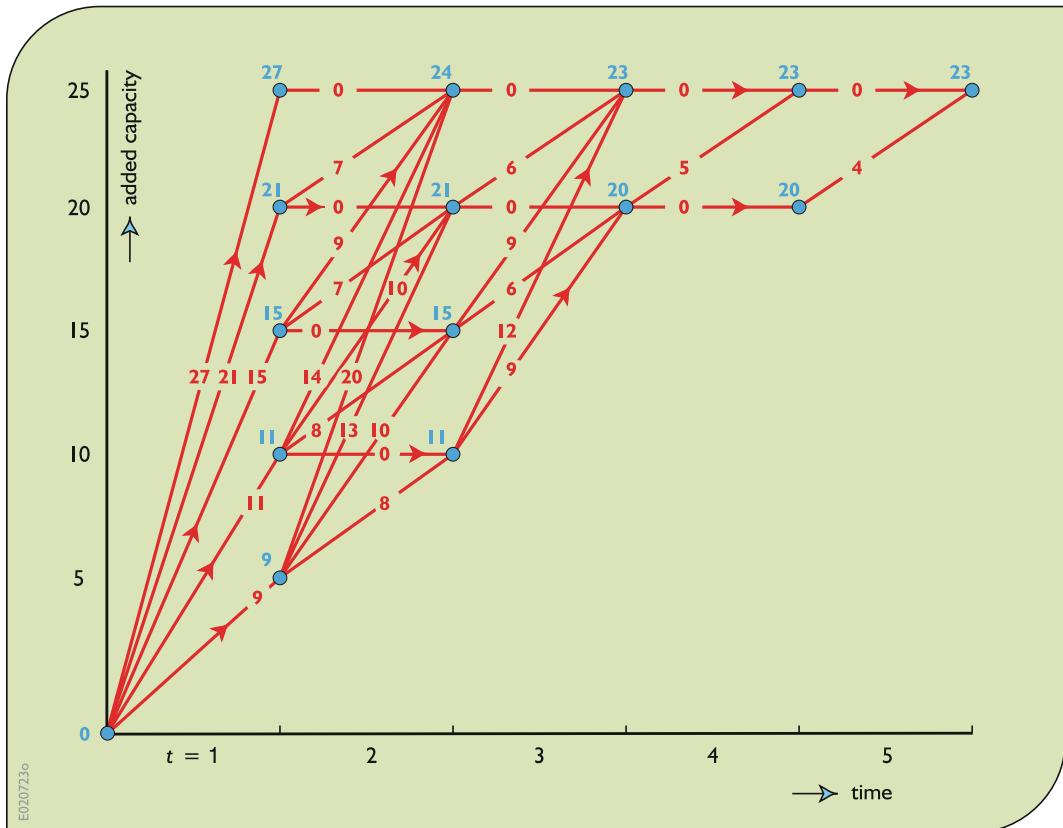
From the forward-moving solution to the dynamic programming problem shown in Fig. 4.14, the present value of the cost of the optimal capacity expansion schedule is 23 units of money. Backtracking (moving left against the arrows) from the farthest right node, this schedule adds 10 units of capacity in 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

Fig. 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

Fig. 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

$$F_t(s_t) = \text{minimum}\{C_t(s_t, x_t) + F_{t+1}(s_{t+1})\}$$
$$\text{over all discrete } x_t \text{ from } D_t - s_t \text{ to } D_T - s_t,$$
$$\text{for all discrete states } s_t \text{ from } D_{t-1} \text{ to } D_T$$

$$(4.106)$$
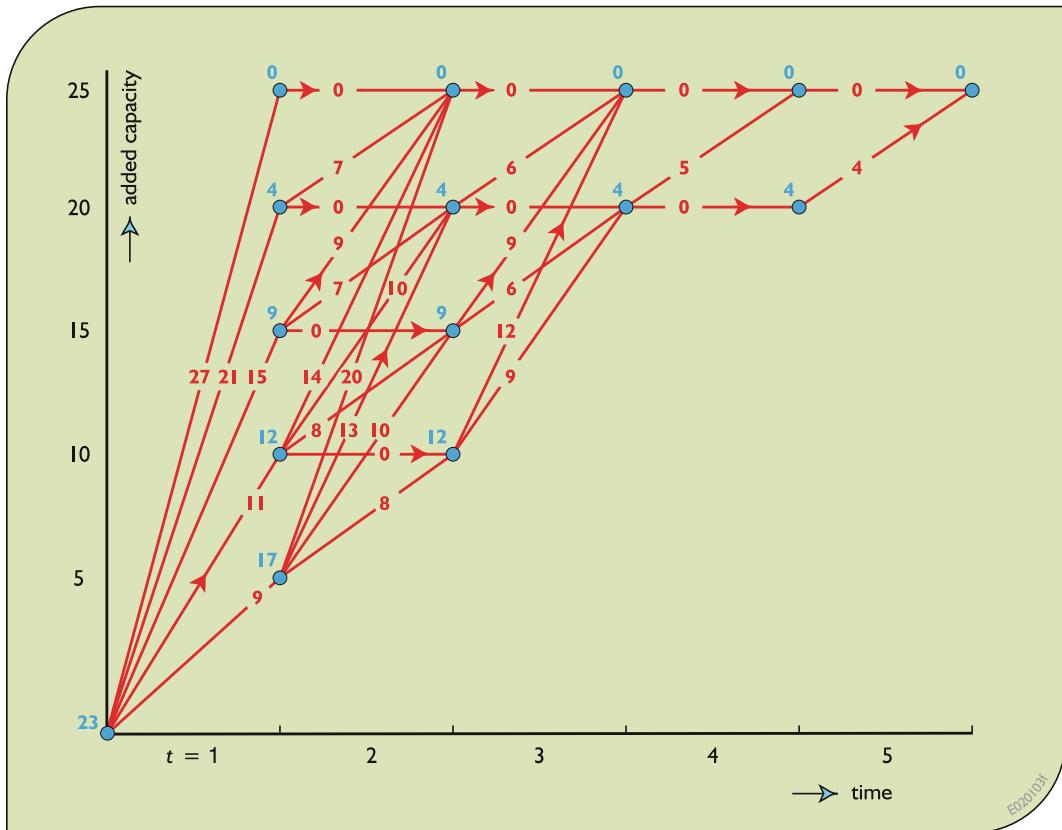
where $F_{T+1}(D_T) = 0$ and as before each cost function is the discounted cost.

Once again, as shown in Fig. 4.14, 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 consider 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 in the current period, i.e., 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 ask. If the answer is no, then one can place some confidence in the value of x_1. If the answer is 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 time periods 2, 3, and so on can be based on updated information and analyses carried out closer to the

Fig. 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

time those decisions are to be made. At those times, the forecast demands and 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 Figs. 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 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 if any 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.

4.4.7.2   Reservoir Operation
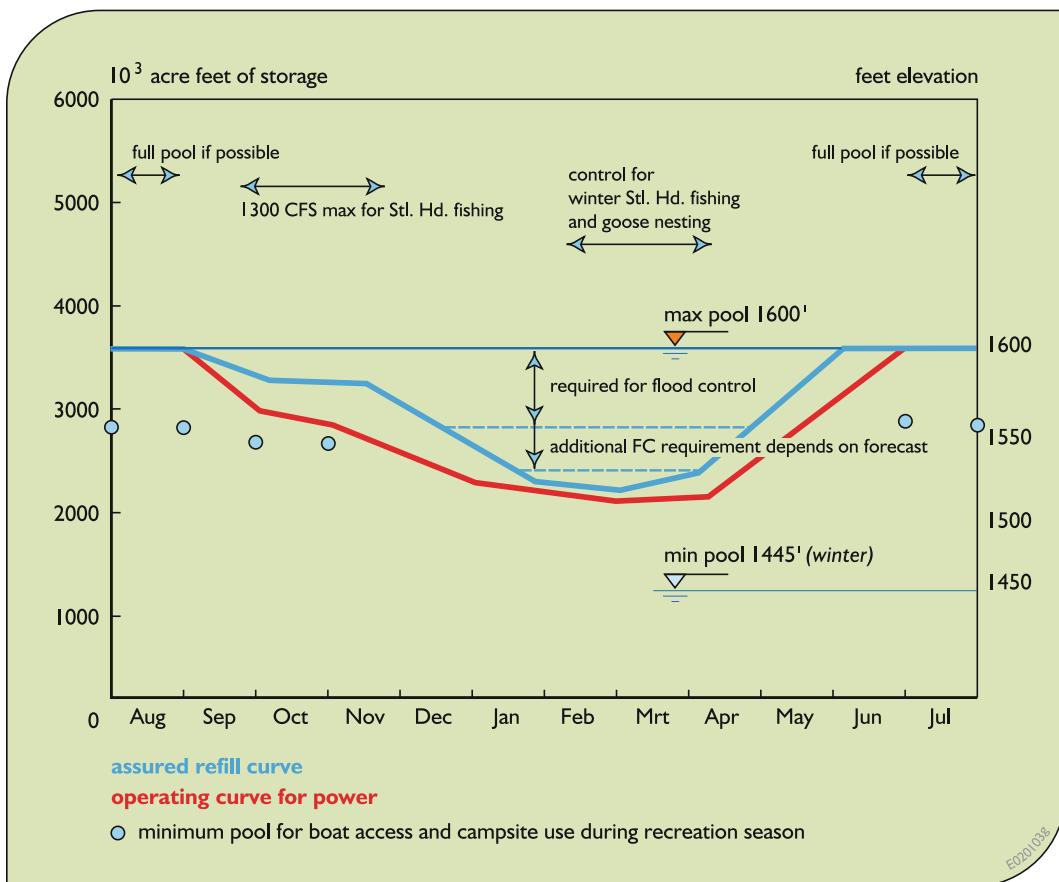
Reservoir operators need to know how much water to release and when. Reservoirs designed to meet demands for water supplies, recreation, hydropower, the environment and/or flood control need to be operated in ways that meet those

demands in a 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 that could exist or happen in each time period t in the future.

Reservoir release policies are often defined in the form of what are called "rule curves." Figure 4.17 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 discrete 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 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.
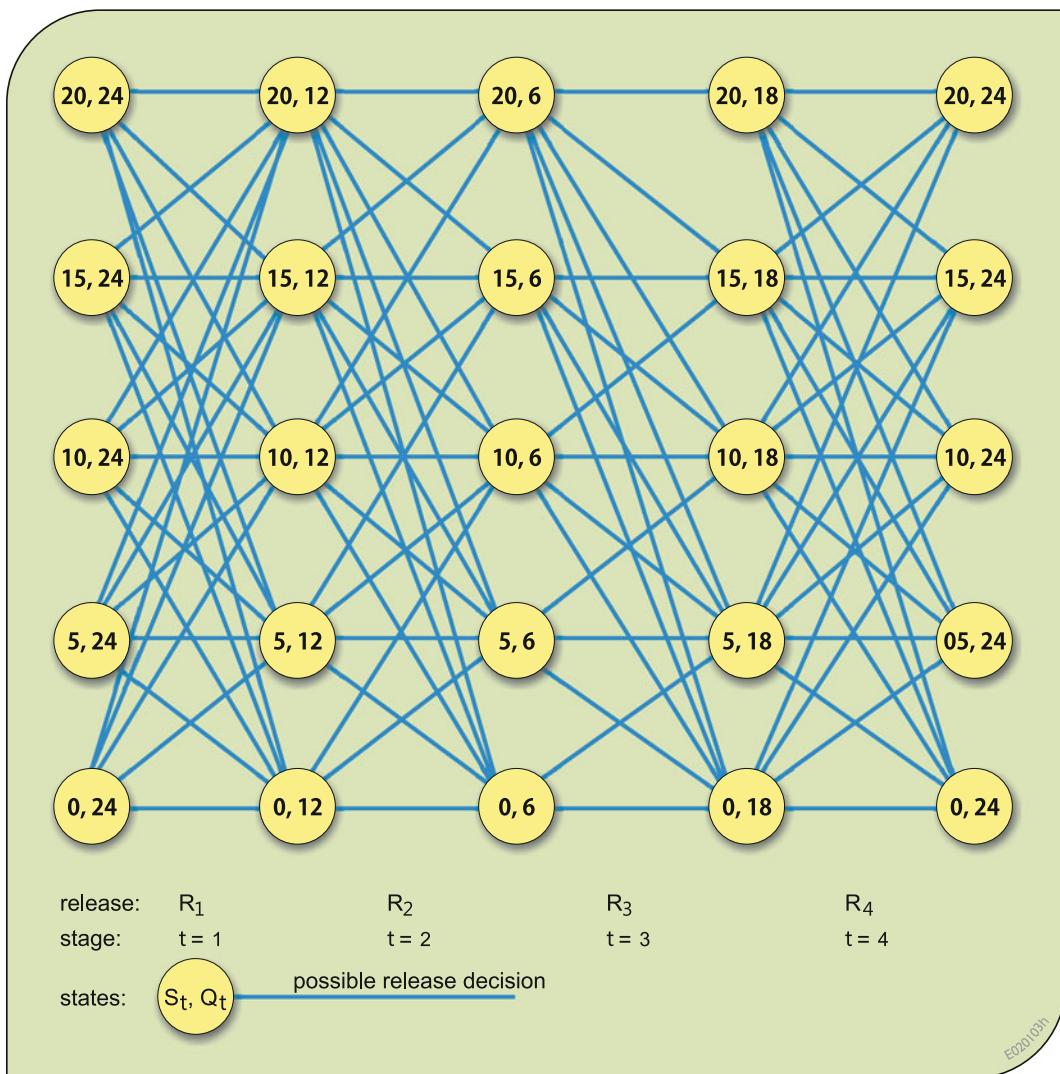
Fig. 4.16 An example reservoir rule curve specifying the storage targets and some of the release constraints, given the particular current storage volume and time of year. The release constraints also include the minimum

and maximum release rates and the maximum downstream channel rate of flow and depth changes that can occur in each month

Reservoir release rule curves for a year, such as that shown in Fig. 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—the two axes of Fig. 4.16.

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

Fig. 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

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 Fig. 4.17.

A numerical example

As a simple example, consider a reservoir having an active storage capacity of 20 million cubic meters, 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 initial and final storage volumes 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

$$S_t + Q_t - R_t - L_t(S_t, S_{t+1}) = S_{t+1} \text{ for each period } t.$$
$$(4.107)$$

To satisfy the requirement (imposed for convenience in this example) that each storage volume variable be a discrete value over the range 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 five 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 Eq. 4.107. 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 Fig. 4.17.

This reservoir-operating problem is a multi-stage decision-making problem. As Fig. 4.17 illustrates, at the beginning of any season t, the storage volume can be in any of the five discrete states. Given the 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 initial and final storage volumes, as defined 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

Table 4.6 Evaporation and seepage losses based on initial and final storage volumes for example reservoir-operating problem

period t = 1

initial storage	final storage 0	5	10	15	20
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

losses

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

losses

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

losses

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 4.7   Discrete releases associated with initial and final storage volumes for example reservoir-operating problem

inflow $Q_1 = 24$, period $t = 1$	initial storage	final storage					releases
		0	5	10	15	20	
	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 $t = 2$	initial storage	final storage					releases
		0	5	10	15	20	
	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 $t = 3$	initial storage	final storage					releases
		0	5	10	15	20	
	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 $t = 4$	initial storage	final storage					releases
		0	5	10	15	20	
	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	

E020820o

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

that the marginal "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 few larger ones.

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} in appropriate periods t.

The objective is to minimize the sum of total weighted squared deviations, TSD_t, over all seasons t from now on into the future:

$$\text{Minimize} \quad \sum_t TSD_t, \qquad (4.108)$$

where

$$TSD_t = ws_t \left[(TS - S_t)^2 + (TS - S_{t+1})^2 \right]$$
$$+ wf_t \left[(ES_t)^2 + (ES_{t+1})^2 \right] + wr_t \left[DR_t^2 \right]$$
$$(4.109)$$

In the above equation, when $t = 4$, the last period of the year, the following period $t + 1 = 1$, the first period in the following year. Each ES_t is the storage volume in excess of the flood storage target volume, TF. 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, above the flood target storage TF at the beginning of each season t can be defined by the constraint:

$$S_t \leq TF + ES_t \quad \text{for periods } t = 1 \text{ and } 2.$$
$$(4.110)$$

The deficit release, DR_t, during period t can be defined by the constraint:

$$R_t \geq TR_t - DR_t \quad \text{for all periods } t. \qquad (4.111)$$

The first component of the right side of Eq. 4.109 defines the weighted squared deviations from a recreation storage target, TS, 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 Eq. 4.109 is for flood control. It defines the weighted squared

deviations associated with storage volumes in excess of the flood control target volume, TF, at the beginning and end of the flood season, period $t = 1$. In this example, the weights, wf_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 Eq. 4.109 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 Fig. 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, 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 Eq. 4.109, 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 or periods 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 multiyear network—each year of which is as shown in Fig. 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 does not 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, toward 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

$$F_t^n(S_t, Q_t) = \min \sum_{t=1,n} \text{TSD}_t(S_t, R_t, S_{t+1})$$

over all feasible values of R_t,

(4.112)

where

$$S_{t+1} = S_t + Q_t - R_t - L_t(S_t, S_{t+1}) \quad (4.113)$$

and

$$S_t \leq K, \text{ the capacity of the reservoir} \quad (4.114)$$

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 and a repeatable hydrologic time series of seasonal inputs. We can find this steady-state policy by first assuming that at some time all future benefits, losses or penalties, $F_t^\circ(S_t, Q_t)$, will be 0.

We can begin in that last season t of reservoir operation and work backwards toward 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.

Table 4.9 Total sum of squared deviations, TSD_t, associated with initial and final storage volumes

$TR_1 \geq 10,\ TS_t^F \leq 15$
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	TDS_1

$TS_2 = 20,\ TR_2 \geq 15$
period t = 2

initial 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	TDS_2

$TS_3 = 20,\ TR_3 \geq 20$
period t = 3

initial 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	TDS_3

$TR_4 \geq 15$
period t = 4

initial 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	TDS_4

E020820q

These are calculated using Eqs. 4.109–4.111

(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 for specific within-year periods 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 Fig. 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, $F_t^{\circ}(S_t, Q_t)$ associated with each state (S_t, Q_t), i.e., each node, equal 0. Since for this problem there is no future. Now we can begin the process of finding the best releases R_t in each successive season t, 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 of remaining seasons n by one.

At each stage, or season t, for each discrete state (S_t, Q_t) we can compute the release R_t or equivalently the final storage volume S_{t+1}, that minimizes

$$F_t^n(S_t, Q_t) = \text{Minimum}\{TSD_t(S_t, R_t, S_{t+1}) + F_{t+1}^{n-1}(S_{t+1}, Q_{t+1})\} \text{ for all } 0 \le S_t \le 20 \tag{4.115}$$

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

$$R_t \le S_t + Q_t - L_t(S_t, S_{t+1}) \tag{4.116}$$

$$R_t \ge S_t + Q_t - L_t(S_t, S_{t+1}) - 20 \text{ (the capacity)} \tag{4.117}$$

and

$$S_{t+1} = S_t + Q_t - R_t - L_t(S_t, S_{t+1}) \tag{4.118}$$

If the decision variable is the final storage volume, S_{t+1}, the constraints on that final storage volume are

$$0 \le S_{t+1} \le 20 \tag{4.119}$$

$$S_{t+1} \le S_t + Q_t - L_t(S_t, S_{t+1}) \tag{4.120}$$

and

$$R_t = S_t + Q_t - S_{t+1} - L_t(S_t, S_{t+1}) \tag{4.121}$$

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, 4.11, 4.12, 4.13, 4.14, 4.15, 4.16, 4.17, 4.18 and 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. This indicates we have reached a steady-state policy. If we kept on computing the release and final storage policies for preceding seasons, we would 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,

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					F_4^1	R_4	S_1
	0	5	10	15	20			
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					$F_3^2(S_3, Q_3)$	R_3	S_4
	0	5	10	15	20			
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

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 = 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					F_2^3 (S_2, Q_2)	R_2	S_3
	0	5	10	15	20			
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 4.13 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 = 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					F_1^4 (S_1, Q_1)	R_1	S_2
	0	5	10	15	20			
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 4.14 Calculation of minimum squared deviations associated with various discrete storage states in season $t = 4$ with $n = 5$ 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					F_4^5 (S_4, Q_4)	R_4	S_1
	0	5	10	15	20			
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

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					F_3^6 (S_3, Q_3)	R_3	S_4
	0	5	10	15	20			
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

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 $t = 2$ $\quad$ $n = 7$ $\quad$ $Q_2 = 12$

$$F_2^7(S_2, Q_2) = \min \{TSD_2(S_2, R_2, S_3) + F_3^6(S_3, Q_3)\}$$

initial storage S_2	$TSD_2 + F_3^6(S_3, Q_3)$ final storage S_3					F_2^7 (S_2, Q_2)	R_2	S_3
	0	5	10	15	20			
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 $t = 1$ $\quad$ $n = 8$ $\quad$ $Q_2 = 24$

$$F_1^8(S_1, Q_1) = \min \{TSD_1(S_1, R_1, S_2) + F_2^7(S_2, Q_2)\}$$

initial storage S_1	$TSD_1 + F_2^7(S_2, Q_2)$ final storage S_2					F_1^8 (S_1, Q_1)	R_1	S_2
	0	5	10	15	20			
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

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 = 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					F_4^9 (S_4, Q_4)	R_4	S_1
	0	**5**	**10**	**15**	**20**			
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 4.19 Calculation of minimum squared deviations associated with various discrete storage states in season $t = 3$ with $n = 10$ 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					F_3^{10} (S_3, Q_3)	R_3	S_4
	0	**5**	**10**	**15**	**20**			
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 4.20 The discrete steady-state reservoir-operating policy as computed for this example problem in Tables 4.16, 4.17, 4.18 and 4.19

| initial storage S	release season t				final storage volume season t			
	1	2	3	4	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	20	15	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
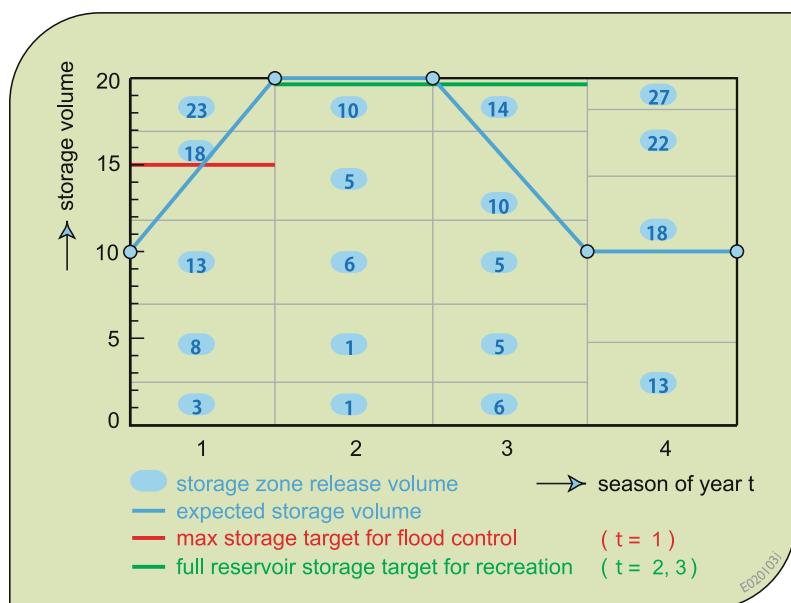
but not on the year. This policy as defined in Tables 4.16, 4.17, 4.18 and 4.19 is summarized in Table 4.20.

This policy can be defined as a rule curve, as shown in Fig. 4.18. It provides a first approximation of a reservoir release rule curve that one can improve upon using simulation.

Table 4.20 and Fig. 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 (Chap. 9).

Fig. 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 equaled the inflows used to derive this rule curve

Table 4.21 A simulation of the derived operating policy in Table 4.20

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	13	1	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

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

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 2 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 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 Fig. 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, 4.13, 4.14, 4.15 and 4.16, 4.17, 4.18, 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 equals a constant, namely the annual value of the objective function. In this case that constant is 186.

$$F_t^{n+1}(S_t, Q_t) - F_t^n(S_t, Q_t) = 186 \quad \text{for all} \quad S_t, Q_t \text{ and } t. \tag{4.122}$$

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 Fig. 4.18 is much more useful than that in Table 4.21. In Chap. 8 we will modify this reservoir operation model to define releases or final storage volumes as functions of not only discrete storage volumes S_t but also of discrete possible inflows Q_t. However, the policy defined by any relatively simple optimization model policy should be simulated, evaluated, and further refined in an effort to identify the policy that best meets the operating policy objectives.

4.4.8 General Comments on Dynamic Programming

Before ending this discussion of using dynamic programming methods for analyzing water resources planning, management and operating policy problems, we should examine a major 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 at each stage of the problem are dependent only on the state and decision 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, then dynamic programming, with some exceptions, becomes more difficult to apply. Dynamic programming models can be applied to design problems, such as the capacity expansion problem 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 at the same time. 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, management and operating policy studies, other optimization methods, perhaps combined with dynamic programming where appropriate, are often more useful.

4.5 Linear Programming

If the objective function and constraints of an optimization model are all linear, many readily available computer programs exist for finding its optimal solution. Surprisingly many water resource systems problems meet these conditions of linearity. These linear optimization programs are very powerful, and unlike many other optimization methods, they can be applied successfully to very large optimization problems containing many variables and constraints. Many water resources problems are too large to be easily solved using nonlinear or dynamic programming methods. The number of variables and constraints simply defining mass balances and capacity limitations in numerous time periods

can become so big as to preclude the practical use of most other optimization methods. Linear programming procedures or algorithms for solving linear optimization models are often the most efficient ways to find solutions to such problems. Hence there is an incentive to convert large optimization models to a linear form. Some ways of doing this are discussed later in this chapter.

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 in practical use today. 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.

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 nonlinear 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 a 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 4.41 and 4.42 define the general structure of any constrained optimization problem. If the objective function $F(X)$ of the vector X of decision variables x_j is linear and if all the constraints $g_i(X)$ in Eq. 4.42 are linear, then the model becomes a linear programming model.

The general structure of a linear programming model is

$$\text{Maximize or minimize} \quad \sum_j P_j x_j \quad (4.123)$$

Subject to

$$\sum_j a_{ij} x_j \le \text{or} \ge b_i \quad \text{for} \ i = 1, 2, 3, \ldots, m \quad (4.124)$$
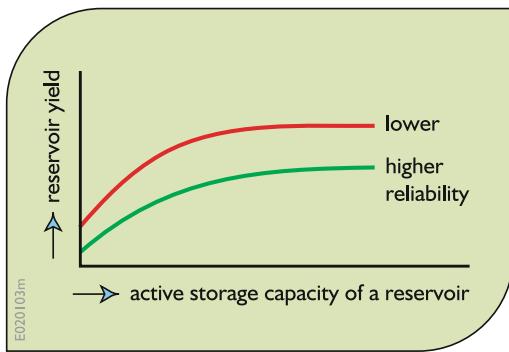
$$x_j \ge 0 \quad \text{for} \ j = 1, 2, 3, \ldots, n. \quad (4.125)$$

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. Variable nonnegativity 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.

Potential users of linear programming algorithms 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.

4.5.1 Reservoir Storage Capacity-Yield Models

Linear programming can be used to define storage capacity-yield functions for a single or

Fig. 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

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, i.e., their inflows. 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 period of record. (In reality, that or any nonzero yield will have a reliability less than 1, as will be considered in Chaps. 6 and 10.)

If a reservoir having an active storage capacity of 1 is built, it could store 1 volume unit of flow when the flow is greater than 2. It could then release it along with the natural flow when the natural flow is 1, increasing the 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 Chap. 6.)

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 Chap. 10.

To illustrate a storage capacity-yield model, consider a single reservoir that must provide at least 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 constant yield Y obtainable from a given active storage capacity. The objective is to

$$\text{maximize } Y \qquad (4.126)$$

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.)

$$S_t + Q_t - Y - R_t = S_{t+1} \text{ for each period } t$$
$$= 1, 2, 3, \ldots, T. \ T + 1 = 1$$
$$(4.127)$$

If, as indicated in Eq. 4.127, 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 as well as the available storage capacity, K.

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.

$$S_t \leq K \quad t = 1, 2, 3, \ldots, T \qquad (4.128)$$

To derive a storage-yield function, the model defined by Eqs. 4.126–4.128 must be solved for various assumed values of capacity K. Only the inflow values Q_t and reservoir active storage capacity K are assumed 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 functions 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 an assumed value of the yield exceeds the mean inflow, there will be no feasible solution to the linear programming model.

Box 4.1 illustrates an example storage-yield model and its solutions to find the storage-yield function. For this problem, and others in this chapter, the program LINGO (freely obtained from www.lindo.com) is used.

Box 4.1. Example storage capacity-yield model and its solution from LINGO

```
! 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

Before moving to another application of linear programming, consider how this storage-yield problem, Eqs. 4.126–4.128, 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 finding a storage-yield function provides a good exercise in dynamic programming. The dynamic programming network has the same form as shown in Fig. 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 at the beginning of 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:

$$F_t^n(S_t) = \max\big\{\min\big[R_t, F_{t+1}^{n-1}(S_{t+1})\big]\big\} \quad (4.129)$$

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

$$\begin{aligned} R_t &\le S_t + Q_t \\ R_t &\ge S_t + Q_t - K, \quad \text{and} \\ S_{t+1} &= S_t + Q_t - R_t \end{aligned} \quad (4.130)$$

These recursive Eqs. 4.129 together with constraint Eqs. 4.130 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

$$\begin{aligned} F_t^n(S_t) &= F_t^{n-1}(S_t) \\ &\text{for all values of } S_t \text{ and periods } t. \end{aligned} \quad (4.131)$$
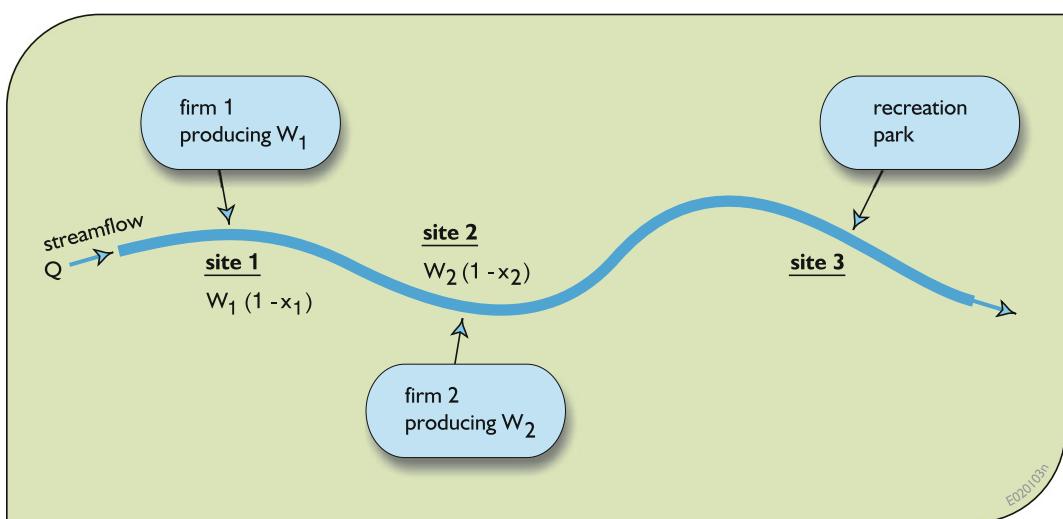
The steady-state yields $F_t(S_t)$ will depend on the storage volumes S_t. High initial storage volumes will result in higher yields than will lower ones. The highest yield will be that associated with the highest storage volumes and it will equal the same value obtained from either of the two linear programming models.

4.5.2   A Water Quality Management Problem

Some linear programming modeling and solution techniques can be demonstrated using the simple water quality management example shown in Fig. 4.21. 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 Fig. 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 just upstream of site 2 and at site 3 at a minimum total cost.

Fig. 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

This is the classic water quality management problem that is frequently found in the literature, although least-cost solutions have rarely if ever 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 as well as 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 W_j as the mass of pollutant generated at site j ($j = 1, 2$) each day. Without any treatment and assuming no upstream pollution concentration, the discharge of W_1 (in units of mass per unit time, (M/T) at site $j = 1$ results in pollutant concentration of P_1 in the stream at that site. This concentration, (M/L^3) equals the discharge W_1 (M/T) divided by the streamflow Q_1 (L^3/T) at that site. For example, assuming the concentration is expressed in units of mg/l and the flow is in terms of m^3/s, and mass of pollutant discharged is expressed as kg/day, and the flow component of the wastewater discharge is negligible compared to the streamflow, the resulting streamflow concentration P_1 at site $j = 1$ is $W_1/86.4\,Q_1$:

$$
\begin{aligned}
P_1(\text{mg/l}) &= \text{Mass } W_1 \text{ discharged at site 1 (kg/day)/} \\
&\quad \text{streamflow } Q_1 \text{ at site } 1\,(\text{m}^3/\text{s})/ \\
&\quad (\text{kg}/10^6\,\text{mg})(86,400\,\text{s/day})(10^3\,\text{L/m}^3) \\
&= W_1/86.4\,Q_1
\end{aligned}
$$

$$(4.132)$$

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 α_{ij} of the mass at site i that reaches site j is often assumed to be

$$\alpha_{ij} = \exp(-kt_{ij}), \qquad (4.133)$$

where k is a rate constant (1/time unit) that depends on the pollutant and the temperature, and t_{ij} is the time (number of time units) it takes a particle of pollutant to flow from site i to site j. 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 decay 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 α_{12} and α_{23}:

$$\alpha_{13} = \alpha_{12}\alpha_{23} \qquad (4.134)$$

In general, for any site k between sites i and j:

$$\alpha_{ij} = \alpha_{ik}\alpha_{kj} \qquad (4.135)$$

Knowing the α_{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 Eq. 4.133. 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 saline 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 conditions are set at low-flow values (e.g., the lowest monthly average flow expected once in twenty years, or the minimum 7-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 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 design low-flow value.

In any event, given particular design streamflow and temperature conditions, our first job is to determine the values of these dimensionless transfer coefficients α_{ij}. They will be independent of the amount of waste discharged into the stream as long as the stream stays aerobic. To determine both α_{12} and α_{23} in this example problem (Fig. 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 at site 1 and ends just upstream of site 2, the mass (kg/day) of pollutants just upstream of site 2 will equal the mass at site 1, W_1, times the transfer coefficient α_{12}:

$$\text{Mass just upstream of site } 2 = W_1\alpha_{12} \quad (4.136)$$

From this equation and 4.132 one can calculate the concentration of pollutants just upstream of site 2.

The mass of additional pollutant discharged into site 2 is W_2. Hence the total mass just downstream of site 2 is $W_1\alpha_{12} + W_2$. At site 3 the pollutant mass will equal the mass just downstream of site 2, times the transfer coefficient a_{23}. Given a streamflow of Q_3 m³/s and pollutant masses W_1 and W_2 kg/day, the pollutant concentration P_3 expressed in mg/l will equal

$$P_3 = [W_1\alpha_{12} + W_2]\,\alpha_{23}/(86.4Q_3) \quad (4.137)$$

4.5.2.1 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 (extending from site 1 to site 2, and from site 2 to site 3) 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 m³/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 for the transfer coefficients α_{12} and α_{23} after adding error terms. More than one sample is needed to allow for measurement errors and other random effects such as those from varying temperature, 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 P_{1s} and P_{2s} and their combined error as E_s. Thus

$$\text{P}_{2s} + E_s = P_{1s}\alpha_{12}(Q_1/Q_2) \quad (4.138)$$

The problem is to find the best estimates of the unknown α_{12}. One way to do this is to define "best" as those values of α_{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

$$\text{Minimize} \quad \sum_s |E_s| \quad (4.139)$$

The set of Eqs. 4.138 and 4.139 is an optimization model. The absolute value signs in Eq. 4.139 can be removed by writing each error term as the difference between two nonnegative variables, $\text{PE}_s - \text{NE}_s$. Thus for each sample pair s:

$$E_s = PE_s - NE_s \qquad (4.140)$$

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 nonnegative. If E_s is positive, it will equal PE_s, and NE_s will be 0. The objective function, Eq. 4.139, 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:

$$\text{Minimize} \quad \sum_s (PE_s + NE_s) \qquad (4.141)$$

Equations 4.139 and 4.140, together with objective function 4.141 and a set of measurements, P_{1s} and P_{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 α_{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. (In this model the measured concentrations are denoted as SP_{js} rather than P_{js}. Again, the program LINGO (www.lindo.com) is used to solve the model).

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 m³/s due to the addition of wastewater flow at site 2.

As shown in Boxes 4.2 and 4.3, the values of the transfer coefficients are $\alpha_{12} = 0.25$ and $\alpha_{23} = 0.60$. Thus from Eq. 4.134, α_{12} $\alpha_{23} = \alpha_{13} = 0.15$.

4.5.2.2 Management Model

Now that these parameter values α_{ij} are known, a water quality management model can be developed. The water quality management problem, illustrated in Fig. 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 end of the two reaches 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 α_{12} that remains at site 2, divided by the streamflow Q_2 at site 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 mass of pollutant entering the treatment plant at site 1. Similarly for site 2. The fraction of waste removal, x_1, at site 1 is to be determined. Hence the concentration of pollutant just upstream of site 2 is

$$P_2 = [P_1Q_1 + W_1(1 - x_1)]\alpha_{12}/Q_2 \qquad (4.142)$$

The terms P_1 and Q_1 are the pollutant concentration (M/L³) and streamflow (L³/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

$$P_3 = [P_2Q_2 + W_2(1 - x_2)]a_{23}/Q_3 \qquad (4.143)$$

Box 4.2. Calibration of water quality model transfer coefficient parameter a_{12}

```
! 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 4.142 and 4.143 will become the predictive portion of the water quality management model. The remaining parts of the model include the restrictions on the pollutant concentrations just upstream of site 2 and at site 3, and limits on the range of values that each waste removal efficiency, x_i, can assume.

$$P_j \leq P_j^{\max} \quad \text{for } j = 2 \text{ and } 3 \qquad (4.144)$$

$$0 \leq x_i \leq 1.0 \quad \text{for } i = 1 \text{ and } 2. \qquad (4.145)$$

Finally, the objective is to minimize the total cost of meeting the stream quality standards $P_2^{\max}$ and $P_3^{\max}$ specified in Eqs. 4.144. Letting $C_i(x_i)$ represent the cost function of wastewater treatment at sites $i = 1$ and 2, the objective can be written:

$$\text{Minimize} \quad C_1(x_1) + C_2(x_2) \qquad (4.146)$$

The complete optimization model consists of Eqs. 4.142–4.146. There are four unknown decision variables, x_1, x_2, P_2, and P_3.

Some of the constraints of this optimization model can be combined to remove the two unknown concentration values, P_2 and P_3. Combining Eqs. 4.142 and 4.144, the concentration just upstream of site 2 must be no greater than $P_2^{\max}$:

$$[P_1 Q_1 + W_1(1 - x_1)]\alpha_{12}/Q_2 \leq P_2^{\max} \quad (4.147)$$

Combining Eqs. 4.143 and 4.144, and using the fraction α_{13} (see Eq. 4.134) to predict the contribution of the pollutant concentration at site 1 on the pollutant concentration at Site 3:

$$\{[P_1 Q_1 + W_1(1 - x_1)]\alpha_{13} + [W_2(1 - x_2)]a_{23}\}/Q_3 \leq P_3^{\max}$$
$$(4.148)$$

Equation 4.148 assumes that each pollutant discharged into the stream can be tracked downstream, independent of the other pollutants in the stream. Alternatively, Eq. 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 modeling approaches give the same results if the parameter values and cost functions are the same.

Box 4.3. Calibration of water quality model transfer coefficient parameter a_{23}

```
! 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
```

Table 4.22 Parameter values selected for the water quality management problem illustrated in Fig. 4.20

parameter		unit	value	remark
flow	Q_1	m³/s	10	flow just upstream of site 1
	Q_2	m³/s	12	flow just upstream of site 2
	Q_3	m³/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

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 Eqs. 4.145–4.148 and substituting the parameter values in place of the parameters, and recalling that kg/day = 86.4 (mg/l)(m³/s):

The water quality constraint at site 2, Eq. 4.147, becomes

$$[(32)(10) + 250{,}000(1 - x_1)/86.4]0.25/12 \leq 20$$

that when simplified is

$$x_1 \geq 0.78. \tag{4.149}$$

The water quality constraint at site 3, Eq. 4.148, becomes

$$\{[(32)(10) + 250{,}000(1 - x_1)/86.4]0.15 \\ + [80{,}000(1 - x_2)/86.4]0.60\}/13 \leq 20$$

that when simplified is

$$x_1 + 1.28x_2 \geq 1.79. \tag{4.150}$$
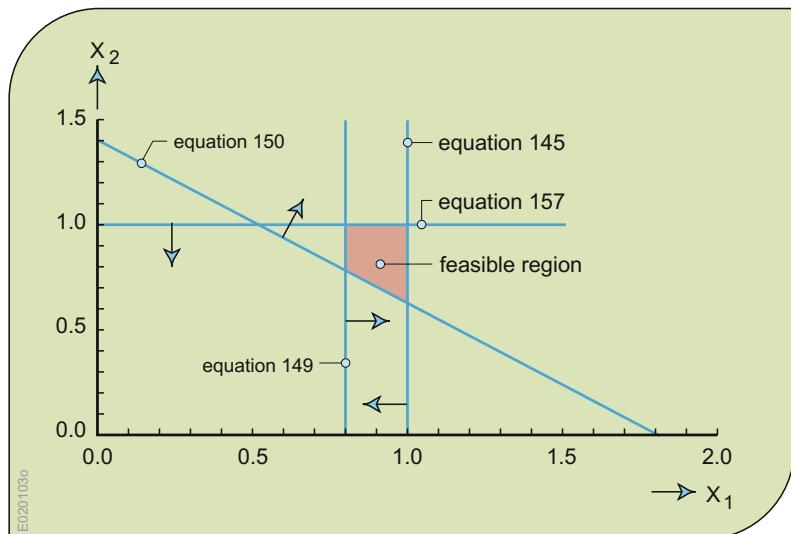
Restrictions on fractions of waste removal, Eq. 4.145, must also be added to this model.

The feasible combinations of x_1 and x_2 can be shown on a graph, as in Fig. 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.
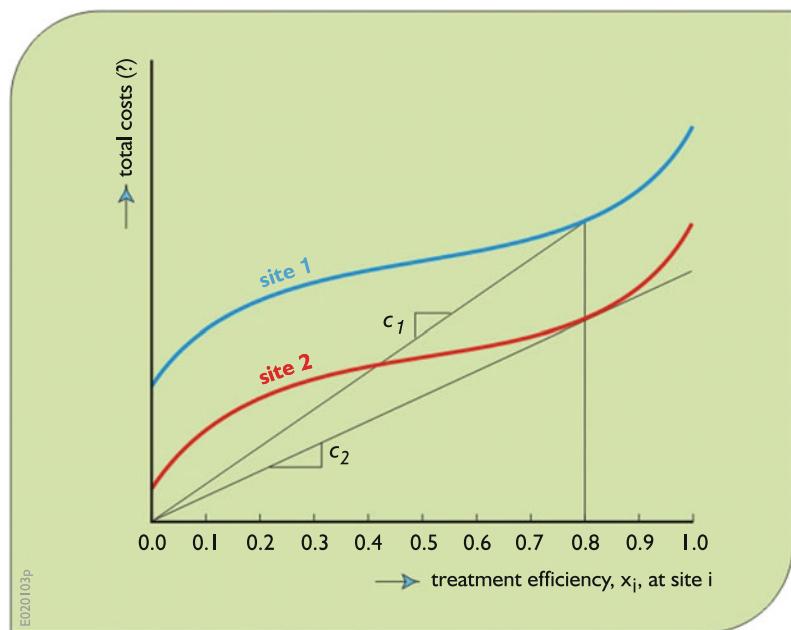
To find the least-cost solution we need the cost functions $C_1(x_1)$ and $C_2(x_2)$ in Eqs. 4.146. Suppose these functions are not known. Can we determine the least-cost solution without knowing these costs? Models like the one just developed can be used to determine just how accurate these cost functions (or the values of any of the model parameters) need to be for the decisions being considered.

While the actual cost functions are not known in this example, their general form can be assumed, as shown in Fig. 4.22. Since the wasteloads produced at site 1 are substantially

Fig. 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. These feasible values are contained in and on the boundaries of the *red* region

Fig. 4.22 General form of total cost functions for wastewater treatment efficiencies at sites 1 and 2 in Fig. 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
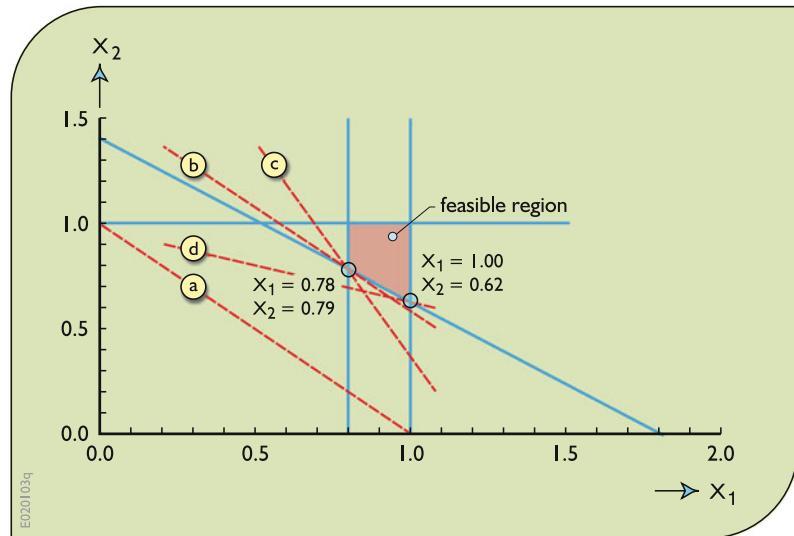
greater than those produced at site 2, and given similar site, labor, and material cost conditions, it seems reasonable to assume that the cost of providing a specified level of treatment at site 1 would exceed (or certainly be no less than) the cost of providing the same specified level of treatment at Site 2. It would also seem the marginal costs at site 1 would be greater than, or at least no less than, the marginal costs at site 2 for any given treatment efficiency. The relative positions of the cost functions shown in Fig. 4.23 are based on these assumptions.

Rewriting the cost function, Eq. 4.146, as a linear function converts the model defined by Eqs. 4.145–4.148 into a linear programming model. For this example problem, the linear programming model can be written as:

Fig. 4.23 Plots of various objective functions (*dashed lines*) together with the constraints of the water quality management model

$$\text{Minimize} \quad c_1 x_1 + c_2 x_2 \qquad (4.151)$$

Equation 4.151 is minimized subject to constraints 4.145, 4.149 and 4.150. 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 (Fig. 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 marginal cost will likely be greater than the other for any particular values of the decision variables x_1 and x_2. In this example that is all we need to know.

First, assume c_1 equals c_2. Let $c_1 x_1 + c_2 x_2$ equal c and assume $c/c_1 = 1$. Thus the cost function is $x_1 + x_2 = 1.0$. This line can be plotted onto the graph in Fig. 4.21, as shown by line "a" in Fig. 4.23.

Line "a" in Fig. 4.23 represents equal values for c_1 and c_2, and the total cost, $c_1 x_1 + c_2 x_2$, 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 also applies for any value of c_1 greater than c_2 (for example line "c" in Fig. 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 Fig. 4.23 and Eq. 4.150, 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 Fig. 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 modeling for purposes other than finding good or "optimal" solutions. Models can help define the necessary precision of the data needed to find those solutions.

Modeling 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 need 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 "nonoptimal" 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 for finding at least an approximate optimal solution. Yet models representing particular water resources systems may not be linear. This motivates the use of methods that can approximate nonlinear functions with linear ones, or the use of other search algorithms such as those discussed in Chap. 5).
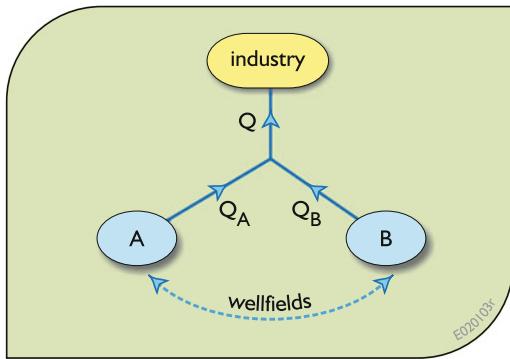
The following simple groundwater supply problem illustrates the application of some linearization methods commonly applied to nonlinear separable functions—functions of only one unknown variable.

These approximation 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.

4.5.3 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 will the water 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 Fig. 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
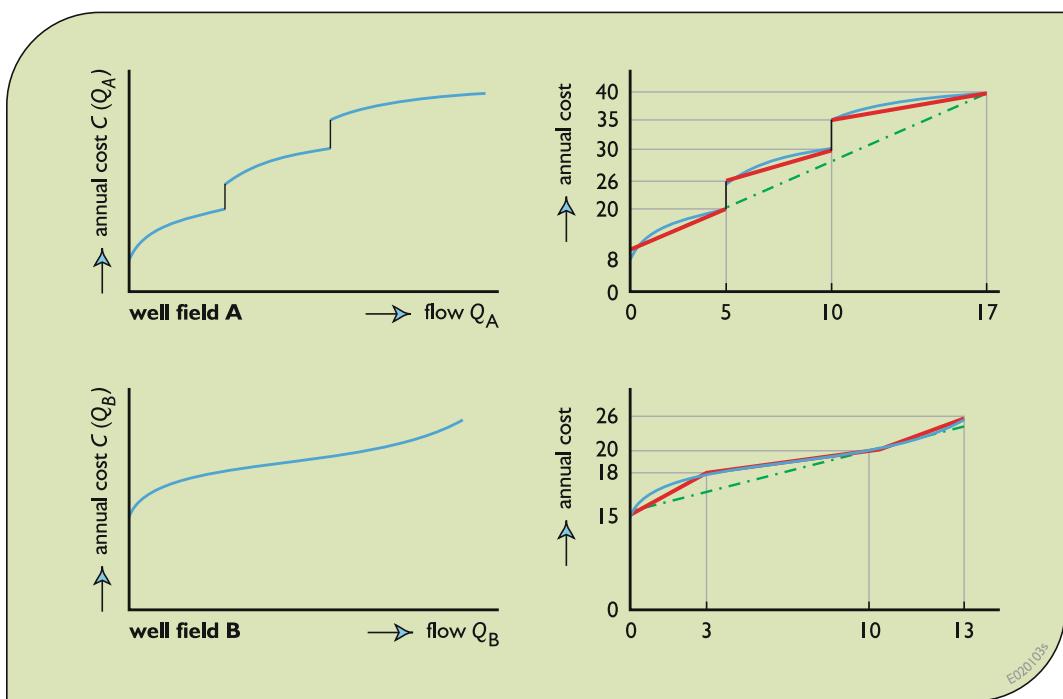
Fig. 4.24 Schematic of a potential groundwater supply system that can serve a water-using industry. The unknown variables are the flows, Q_A and Q_B, from each wellfield

various capacities Q_A and Q_B for Wellfields A and B, respectively, are as shown in Fig. 4.25. These are nonlinear 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 Fig. 4.25, the nonlinear functions on the left have been approximated by piecewise linear functions on the right. This is a first step in linearizing nonlinear separable functions. Increasing the number of linear segments can reduce the difference between the piecewise linear approximation of the actual nonlinear function and the function itself. At the same time it will increase the number of variables and possibly constraints.

When approximating a nonlinear 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 nonlinear) 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.

Fig. 4.25 Annual cost functions associated with the Wellfields A and B as shown in Fig. 4.24. The actual functions are shown on the *left*, and two sets of piecewise linear approximations are shown on the *right*

4.5.3.1 A Simplified Model

Two sets of approximations are shown in Fig. 4.26. Consider first the approximations represented by the light blue dot-dash lines. These single straight lines are very crude approximations of each function. In this example these straight-line cost functions are lower bounds of the actual nonlinear 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 Fig. 4.26, the linear programming model can be written as follows:

$$\text{Minimize} \quad \text{Cost}A + \text{Cost}B \quad (4.152)$$

Subject to

$$\text{Cost}A = 8I_A + [(40-8)/17]Q_A$$
linear approximation of $C(Q_A)$
$$(4.153)$$

$$\text{Cost}B = 15I_B + [(26-15)/13]Q_B$$
linear approximation of $C(Q_B)$
$$(4.154)$$

I_A, I_B are $0, 1$ integer (binary) variables
$$(4.155)$$

$Q_A \leq 17I_A$ limits Q_A to 17 and forces $I_A = 1$
 if $Q_A > 0$
$$(4.156)$$

$Q_B \leq 13I_B$ limits Q_B to 13 and forces $I_B = 1$
 if $Q_B > 0$

$$(4.157)$$

$$Q_A + Q_B = Q \text{ mass balance} \quad (4.158)$$

$Q, Q_A, Q_B \geq 0$
 non-negativity of all decision variables
$$(4.159)$$

$Q =$ some specified amount from 0 to 30.
$$(4.160)$$

The expressions within the square brackets, [ ], in Eqs. 4.154 and 4.155 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 linear 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 Fig. 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.

Fig. 4.26 Least-cost wellfield use given total demand Q based on model defined by Eqs. 4.152 to 4.160
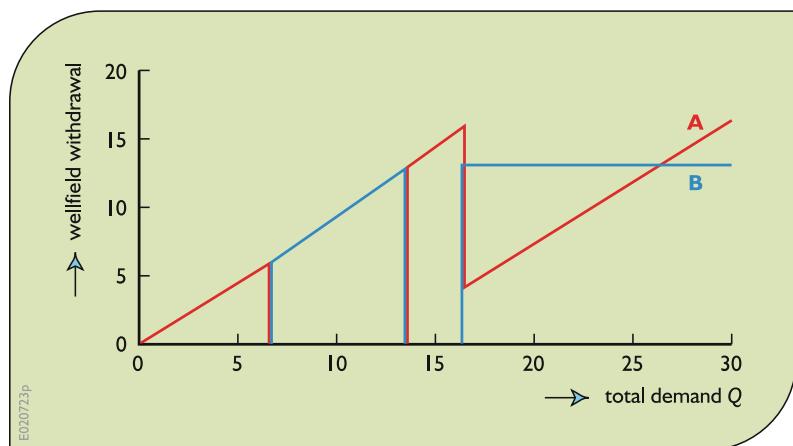
Figure 4.26 illustrates these solutions. One can understand why in situations of increasing demands for Q over time, capacity expansion modeling might be useful. One would not close down a wellfield once developed, just to achieve what would have been a least-cost solution if the existing wellfield had not been developed.

4.5.3.2   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 Fig. 4.26.

$$\text{Minimize} \quad \text{Cost}A + \text{Cost}B \quad (4.161)$$

Subject to

linear approximation of cost functions:

$$\begin{aligned}
\text{Cost}A = &\{8I_{A1} + [(20 - 8)/5]Q_{A1}\} \\
&+ \{26I_{A2} + [(30 - 26)/(10 - 5)]Q_{A2}\} \\
&+ \{35I_{A3} + [(40 - 35)/(17 - 10)]Q_{A3}\}
\end{aligned}$$
$$(4.162)$$

$$\begin{aligned}
\text{Cost}B = &\{15I_{B1} + [(18 - 15)/3]Q_{B1}\} \\
&+ \{18I_{B2} + [(20 - 18)/(10 - 3)]Q_{B2} \\
&+ [(26 - 20)/(13 - 10)]Q_{B3}\}
\end{aligned}$$
$$(4.163)$$

Q_A and Q_B defined.

$$Q_A = Q_{A1} + (5I_{A2} + Q_{A2}) + (10I_{A3} + Q_{A3})$$
$$(4.164)$$

$$Q_B = Q_{B1} + (3I_{B2} + Q_{B2} + Q_{B3}) \quad (4.165)$$

I_{Ai} and I_{Bi} are 0, 1 integer variables
for all segments i $\quad (4.166)$

$$\begin{aligned}
&Q_{A1} \leq 5I_{A1} \\
&Q_{A2} \leq (10 - 5)I_{A2}, \\
&Q_{A3} \leq (17 - 10)I_{A3} \quad \text{limits } Q_{Ai} \text{ to width of}
\end{aligned}$$
segment i and forces $I_{Ai} = 1$ if $Q_{Ai} > 0$
$$(4.167)$$

$$I_{A1} + I_{A2} + I_{A3} \leq 1 \quad \text{limits solution to at most}$$
only one cost function segment i.
$$(4.168)$$

$$\begin{aligned}
&Q_{B1} \leq 3I_{B1}, \\
&Q_{B2} \leq (10 - 3)I_{B2}, \\
&Q_{B3} \leq (13 - 10)I_{B2} \quad \text{limits } Q_{Bi} \text{ to width of}
\end{aligned}$$
segment i and forces $I_{Bi} = 1$ if $Q_{Bi} > 0$.
$$(4.169)$$

$$I_{B1} + I_{B2} \leq 1 \quad (4.170)$$

$$Q = Q_A + Q_B \text{ mass balance} \quad (4.171)$$
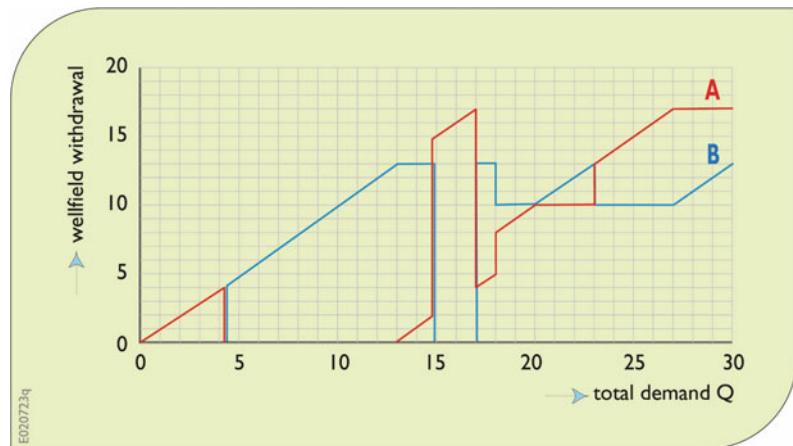
$$Q, Q_A, Q_B \geq 0$$
non-negativity of all decision variables
$$(4.172)$$

$$Q = \text{some specified amount from 0 to 30}$$
$$(4.173)$$

Constraint (4.170) limits the solution to at most only the first segment or to the second and third segments of the cost function for wellfield B. 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 Eqs. 4.169.

The solution to this model, shown in Fig. 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 in excess of 4.3 up to 13 all the water comes from Wellfield B. For values of Q in excess of 13 up to 14.8, the capacity of Wellfield B remains at its maximum capacity of 13 and Wellfield A provides the additional amount of needed capacity over 13. As Q increases from 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 maximum capacity, and the capacity of A increases from 4 to 5. For values of

Fig. 4.27 Least-cost wellfield use given total demand Q based on Eqs. 4.161 to 4.173

Q from 18.1 to 20, Wellfield B decreases to a constant 10, and Wellfield A increases from 8.1 to 10. For values of Q from 20 to 23, Wellfield A remains at 10 and Wellfield B increases from 10 to 13. For values of Q from 23 to 27, Wellfield B again drops to a constant 10 and Wellfield A increases from 13 to 17. For values of Q in excess of 27, Wellfield A remains at its maximum capacity of 17, 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.

4.5.3.3 An Extended Model
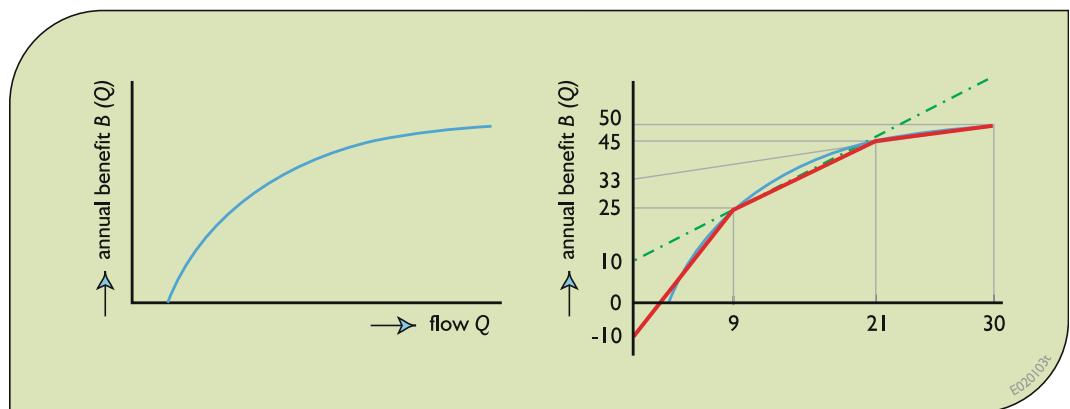
In this example, the simpler model (Eqs. 4.152–4.160) and the more accurate model (Eqs. 4.161–4.173) provided essentially the same allocations of wellfield capacities associated with a specified total capacity Q. If the problem contained a larger number of wellfields, the simpler (and smaller) model 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. It does not address the problem of planning for an increasing demand for Q over time. Clearly it makes no sense to implement the particular cost-effective solution for any value of Q, as shown in Fig. 4.27, as the demand for Q increases, as in this example, from 0 to 30. This is the capacity expansion problem, the solution of which will benefit from models that take time into account and that are not static as illustrated previously in this chapter.

Next, consider a cost–benefit analysis in which the question is just how much water should users 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 Fig. 4.28.

The straight, blue, dot-dash linear approximation of the benefit function shown in Fig. 4.28 is an 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 Fig. 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.

Fig. 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*

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 Eqs. 4.152–4.160 can now be modified. Equation 4.160 is eliminated and the cost minimization objective Eq. 4.152 is replaced with:

$$\text{Maximize} \quad \text{Benefits} - (\text{Cost}A + \text{Cost}B) \tag{4.174}$$

where

$$\begin{aligned} \text{Benefits} &= 10 + [(45 - 25)/(21 - 9)]Q \\ &\text{linear approximation of } B(Q) \end{aligned} \tag{4.175}$$

The solution of this model, Eqs. 4.153–4.159, 4.174, and 4.175 (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 modeling 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 Figs. 4.25 and 4.28.

Note that the approximation of the generally concave benefit function in Fig. 4.29 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 Fig. 4.28. Modeling these initial fixed benefits the same way as the fixed costs have been modeled, 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.

4.5.3.4   Piecewise Linear Model

There are a number of ways of modeling the piecewise linear concave benefit function shown on the right side of Fig. 4.28. Several are defined in the next several sets of equations. Each method will result in the same model solution.

One approach to modeling 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:

$$\text{Benefits} \leq -10 + [(25 - (-10))/9]Q \quad (4.176)$$

$$\text{Benefits} \leq 10 + [(45 - 25)/(21 - 9)]Q \quad (4.177)$$

$$\text{Benefits} \leq 33 + [(50 - 45)/(30 - 21)]Q \quad (4.178)$$

Since most linear programming algorithms assume the unknown variables are nonnegative (unless otherwise specified), unrestricted variables, such as Benefits, can be replaced by the difference between two nonnegative 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 Eqs. 4.176–4.178, and those below, one can substitute Pben − Nben.

Another modeling 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 Fig. 4.28:

$$q_1 \leq 9 \quad (4.179)$$

$$q_2 \leq 21 - 9 \quad (4.180)$$

$$q_3 \leq 30 - 21 \quad (4.181)$$

and

$$Q = q_1 + q_2 + q_3 \quad (4.182)$$

The linearized benefit function can now be written as the sum over all three segments of each segment slope times the variable q_i:

$$\begin{aligned} \text{Benefits} = &-10 + [(25 + 10)/9]q_1 \\ &+ [(45 - 25)/(21 - 9)]q_2 \\ &+ [(50 - 45)/(30 - 21)]q_3 \end{aligned}$$
$$(4.183)$$

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 Eqs. 4.179–4.181.

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 unknown weights must also sum to 1. Hence, for this example:

$$\text{Benefits} = (-10)w_1 + 25w_2 + 45w_3 + 50w_4 \quad (4.184)$$

$$Q = 0w_1 + 9w_2 + 21w_3 + 30w_4 \quad (4.185)$$

$$1 = w_1 + w_2 + w_3 + w_4 \quad (4.186)$$

For this method to provide the closest approximation of the original nonlinear function, the solution must include no more than two nonzero weights and those nonzero weights must be adjacent to each other. For concave functions that are 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 Eqs. 4.174, 4.162–4.172, and either 4.176–4.178, 4.179–4.183, or 4.184–4.186,

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.

4.5.4   A Review of Linearization Methods

This section reviews the piecewise linearization methods just described and some other approaches for incorporating nonlinear conditions into linear programming models. All of these methods maintain linearity.

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. All variables (X, Y) are assumed nonnegative. All variables Z are assumed to be a binary (0, 1) variables. Note the constraints are all linear.

(a) If $Y \leq 50$ then $X \leq 10$, else $X \geq 15$.
Define constraints:

$$Y \leq 50Z + Y_{max}(1 - Z)$$
$$Y \geq 50(1 - Z)$$
$$X \leq 10Z + X_{max}(1 - Z)$$
$$X \geq 15(1 - Z)$$

(b) If $Y \leq 50$ then $X \leq Y$, else $X \geq Y$.
Define constraints:

$$Y \geq 50Z$$
$$Y \leq 50(1 - Z) + Y_{max}Z$$
$$X \leq Y + X_{max}Z$$
$$X \geq Y - Y_{max}(1 - Z)$$

(c) If $Y \leq 20$ or $Y \geq 80$ then $X = 5$, else $X \geq 10$.
Define constraints:

$$Y \leq 20Z_1 + 80Z_2 + Y_{max}(1 - Z_1 - Z_2)$$
$$Y \geq 20Z_2 + 80(1 - Z_1 - Z_2)$$
$$Z_1 + Z_2 \leq 1$$
$$X \leq 5(Z_1 + (1 - Z_1 - Z_2)) + X_{max}Z_2$$
$$X \geq 5(Z_1 + (1 - Z_1 - Z_2))$$
$$X \geq 10Z_2$$

(d) If $20 \leq Y \leq 50$ or $60 \leq Y \leq 80$, then $X \leq 5$, else $X \geq 10$.
Define constraints:

$$Y \leq 20Z_1 + 50Z_2 + 60Z_3 + 80Z_4$$
$$\quad + Y_{max}(1 - Z_1 - Z_2 - Z_3 - Z_4).$$
$$Y \geq 20Z_2 + 50Z_3 + 60Z_4$$
$$\quad + 80(1 - Z_1 - Z_2 - Z_3 - Z_4)$$
$$Z_1 + Z_2 + Z_3 + Z_4 \leq 1$$
$$X \leq 5(Z_2 + Z_4) + X_{max}(1 - Z_2 - Z_4)$$
$$X \geq 10((Z_1 + Z_3) + (1 - Z_1 - Z_2 - Z_3 - Z_4))$$

Minimizing the absolute value of the difference between two unknown nonnegative variables:

Minimize $|X - Y|$ is equivalent to

$$\text{Minimize } D$$

subject to

$$X - Y \leq D;$$
$$Y - X \leq D;$$
$$X, Y, D \geq 0.$$

or

$$\text{Minimize } (PD + ND)$$

subject to

$$X - Y = PD - ND;$$
$$PD, ND, X, Y \geq 0.$$

Minimizing the maximum or maximizing the minimum

Let the set of variables be $\{X_1, X_2, X_3, \ldots, X_n\}$, Minimizing the maximum of $\{X_1, X_2, X_3, \ldots, X_n\}$ is equivalent to

$$\text{Minimize } U$$

subject to

$$U \geq X_j, \quad j = 1, 2, 3, \ldots, n.$$

Maximizing the minimum of $\{X_1, X_2, X_3, \ldots, X_n\}$ is equivalent to
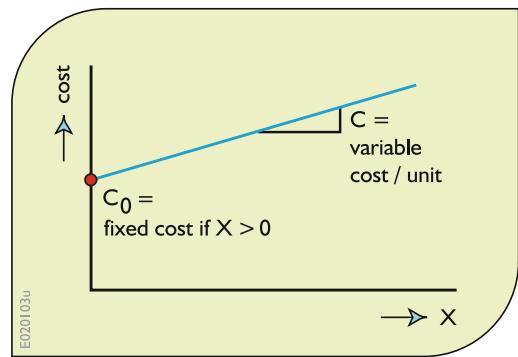
$$\text{Maximize } L$$

subject to

$$L \leq X_j, \quad j = 1, 2, 3, \ldots, n.$$

Linearization of convex functions for maximization or concave function for minimization involves 0, 1 binary variables.

Fixed costs in cost functions

Consider functions that have fixed components if the argument of the function is greater than 0.

$$\begin{aligned}\text{Cost} &= C_0 + CX \quad \text{if } X > 0, \\ &= 0 \qquad \text{otherwise.}\end{aligned}$$

To include these fixed costs in a LP model, define
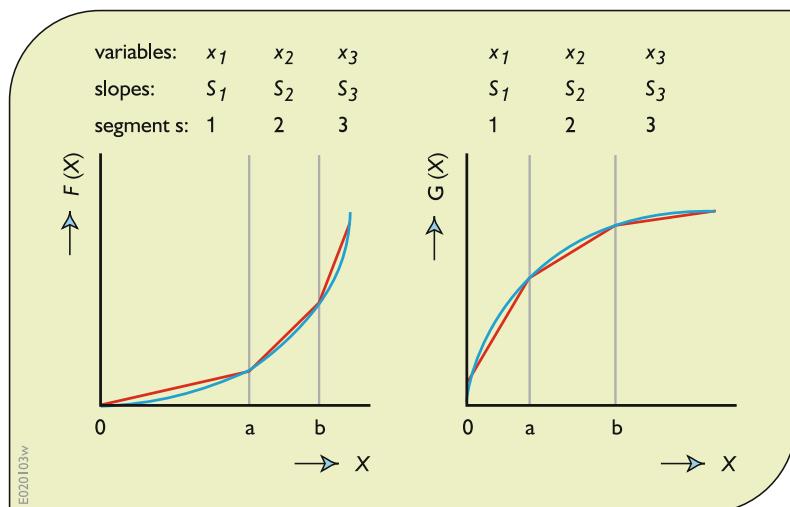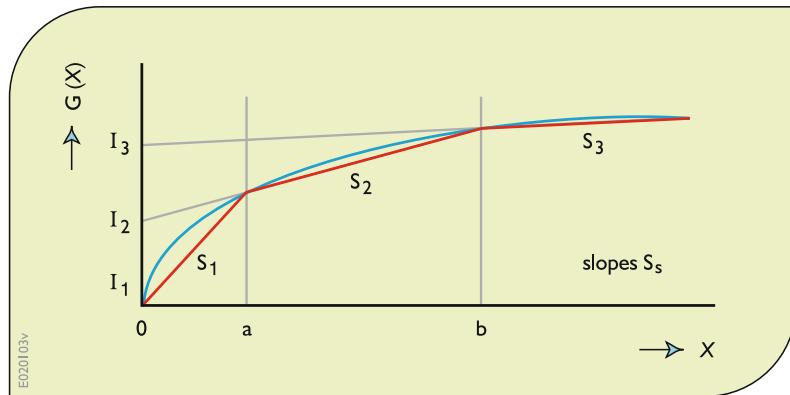
$$\text{Cost} = C_0 I + CX$$
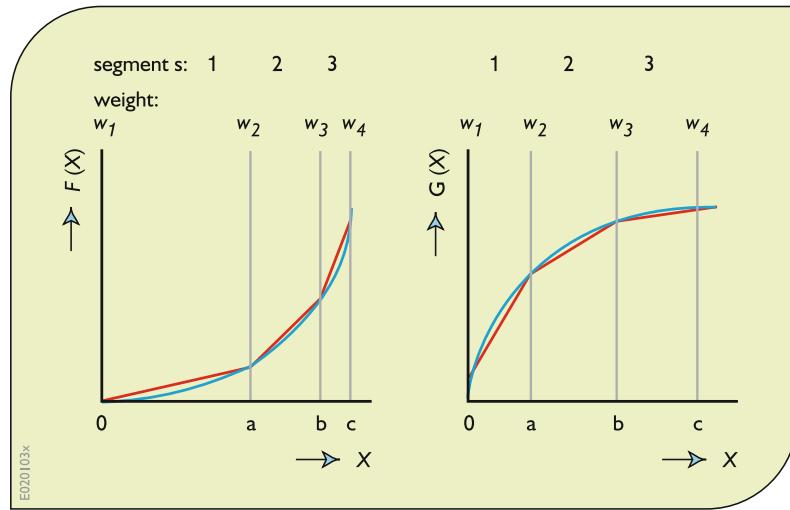
Subject to

$$X \leq MI$$

where M is the maximum value of X, and I is an unknown 0, 1 binary variable.

Minimizing convex functions or maximizing concave functions.

Maximize $G(X)$ = Maximize B

Subject to

$$I_1 + S_1 X \geq B$$
$$I_2 + S_2 X \geq B$$
$$I_3 + S_3 X \geq B$$

Subject to

$$X = x_1 + x_2 + x_3$$
$$x_1 \leq a$$
$$x_2 \leq b - a$$

Minimize $F(X) = S_1 x_1 + S_2 x_2 + S_3 x_3$

Maximize $G(X) = S_1 x_1 + S_2 x_2 + S_3 x_3$

Minimize $F(X) = F(0)w_1 + F(a)w_2 + F(b)w_3 + F(c)w_4$

Maximize $G(X) = G(0)w_1 + G(a)w_2 + G(b)w_3 + G(c)w_4$

Subject to

$$X = 0w_1 + aw_2 + b\,w_3 + c\,w_4$$
$$w_1 + w_2 + w_3 + w_4 = 1$$

Minimizing concave functions or maximizing convex functions

Minimize $G(X) \cong 5x_1 + (20z_2 + 3x_2) + (44z_3 + 2x_3)$

Subject to

$$x_1 + (4z_2 + x_2) + (12z_3 + x_3) = X$$
$$z_s = 0 \text{ or } 1 \text{ for all segments s}$$
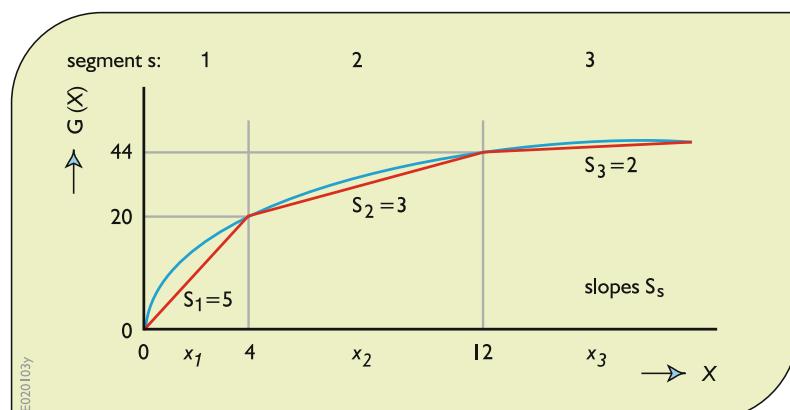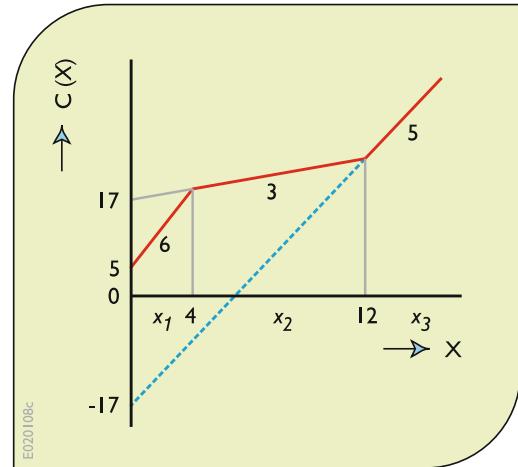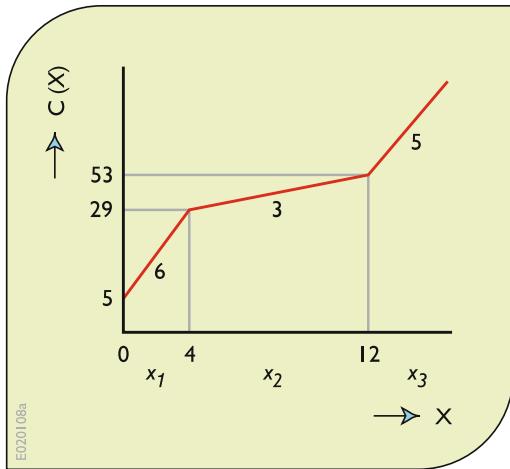
$$x_1 \leq 4z_1;$$
$$x_2 \leq 8z_2;$$
$$x_3 \leq 99z_3;$$
$$z_1 + z_2 + z_3 = 1.$$

Minimizing or maximizing combined concave–convex functions

Maximize
$$C(X) = (5z_1 + 6x_1 + 3x_2) + (53z_3 + 5x_3)$$

Subject to

$$(x_1 + x_2) + (12z_3 + x_3) = X$$
$$x_1 \leq 4z_1$$
$$x_2 \leq 8z_1$$
$$x_3 \leq 99z_3$$
$$z_1 + z_3 = 1$$
$$z_1, z_3 = 0, 1$$

Minimize
$$C(X) = (5z_1 + 6x_1) + (29z_2 + 3x_2 + 5x_3)$$

Subject to

$$z_1, z_2 = 0, 1.$$
$$x_1 + (4z_2 + x_2 + x_3) = X$$
$$x_1 \leq 4z_1$$
$$x_2 \leq 8z_2$$
$$x_3 \leq 99z_2$$
$$z_1 + z_2 \leq 1$$

Maximize
$$C(X) = (5z_1 + 6x_1 + 3x_2) + (-17z_3 + 5x_3)$$

Subject to

$$(x_1 + x_2) + x_3 = X$$
$$z_1, z_3 = 0, 1$$
$$x_1 \leq 4z_1$$
$$x_2 \leq 8z_1$$
$$x_3 \leq 99z_3$$
$$z_1 + z_3 = 1;$$

Minimize
$$C(X) = (5z_1 + 6x_1) + (17z_2 + 3x_2 + 5x_3)$$

Subject to
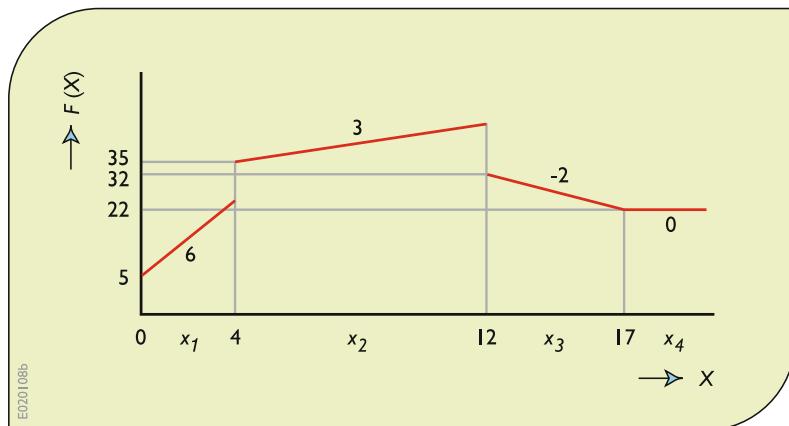
$$x_1 + (4z_2 + x_2 + x_3) = X$$
$$z_1, z_2 = 0, 1.$$
$$x_1 \leq 4z_1$$
$$x_2 \leq 12z_2$$
$$x_3 \leq 99z_2$$
$$z_1 + z_2 \leq 1$$

Maximize or Minimize $F(X)$

$$F(X) = (5z_1 + 6x_1) + (35z_2 + 3x_2) + (32z_3 - 2x_3) + 22z_4$$

Subject to

$$x_1 + (4z_2 + x_2) + (12z_3 + x_3) + (17z_4 + x_4) = X$$
$$x_1 \leq 4z_1$$
$$x_2 \leq 8z_2$$
$$x_3 \leq 5z_3$$
$$x_4 \leq 99z_4$$
$$z_1 + z_2 + z_3 + z_4 = 1;$$
$$z_s = 0, 1 \quad \text{for all segments s}$$

4.6 A Brief Review

Before proceeding to 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 judgment 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. For example, do we minimize the costs of meeting specified maximum levels of pollutant concentrations or minimize pollutant concentrations without exceeding specified costs?

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 analyzed 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 hill-climbing, calculus-based Lagrange multiplier, numerical nonlinear programming, discrete dynamic programming, and linear programming methods. Each method has its strengths and

limitations. Both linear and nonlinear programming models are typically solved using software packages. Many of these software programs are free and readily available. But before any model can be solved, it has to be built. Building models is an art and that is what this chapter has attempted to introduce.

The example problems used to illustrate these modeling 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 following chapters.

Reference

Bellman, R. E. (1957). *Dynamic programming*. Princeton, NJ: Princeton University Press.

Additional References (Further Reading)

Bassson, M. S., Allen, R. B., Pegram, G. G. S., & Van Rooyen, J. A. (1994). Probabilistic management of water resource and hydropower systems. Highlands Ranch, CO: Water Resources Publications.

Beroggi, G. E. G. (1999). Decision modeling in policy management: An introduction to analytic concepts. Boston, MA: Kluwer Academic.

Blanchard, B. S., & Fabrycky, W. J. (1998). *Systems engineering and analysis* (3rd ed.). Prentice Hall: Upper Saddle River, NJ.

Buras, N. (1966). *Dynamic programming and water resources development, advances in hydroscience* (Vol. 3). New York: Academic Press.

Dorfman, R., Jacoby, H. D., & Thomas, H. A., Jr. (1972). Models for managing regional water quality. Cambridge, MA: Harvard University Press.

Esogbue, A. O. (Ed.). (1989). *Dynamic programming for optimal water resources systems analysis*. Prentice Hall: Englewood Cliffs, NJ.

Hall, W. A., & Dracup, J. A. (1970). *Water resources systems engineering*. New York: McGraw-Hill.

Hillier, F. S., & Lieberman, G. J. (1990a). *Introduction to operations research* (5th ed.). New York: McGraw-Hill.

Hillier, F. S., & Lieberman, G. J. (1990b). *Introduction to stochastic models in operations research*. New York: McGraw-Hill.

Hufschmidt, M. M., & Fiering, M. B. (1966). Simulation techniques for design of water-resource systems. Cambridge, MA: Harvard University Press.

Kapamoyz, M., Szidarovszky, F., & Zahpaie, B. (2003). Water resources systems analysis. Boca Raton, FL: Lewis.

Loucks, D. P., & da Costa, J. R. (Eds.). (1990). *Decision support systems: water resources planning*. Berlin: Springer.

Loucks, D. P., Stedinger, J. S., & Haith, D. A. (1981). *Water resource systems planning and analysis*. Prentice Hall: Englewood Cliffs, NJ.

Maass, A., Hufschmidt, M. M., Dorfman, R., Thomas, H. A., Jr., Marglin, S. A., & Fair, G. M. (1962). Design of water-resource systems. Cambridge, MA: Harvard University Press.

Major, D. C., & Lenton, R. L. (1979). *Applied water resources systems planning*. Prentice Hall: Englewood Cliffs, NJ.

Mays, L. W., & Tung, Y.-K. (1992). *Hydrosystems engineering and management*. New York: McGraw-Hill.

Newnan, D. G., Lavelle, J. P., & Eschenbach, T. G. (2002). *Essentials of engineering economic analysis* (2nd ed.). New York: Oxford University Press.

Rardin, R. L. (1998). *Optimization in operations research*. Prentice Hall: Upper Saddle River, NJ.

Revelle, C. (1999). *Optimizing reservoir resources*. New York: Wiley.

Somlyody, L. (1997). Use of optimization models in river basin water quality planning: WATERMATEX'97. In M. B. Beck & 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 ed.). Prentice Hall: Upper Saddle River, NJ.

Willis, R., & Yeh, W. W.-G. (1987). *Groundwater systems planning and management*. Prentice Hall: Englewood Cliffs, NJ.

Winston, W. L. (1994). *Operations research: Applications and algorithms* (3rd edn.). Boston, MA: Duxbury.

Winston, W. L., & Albright, S. C. (2001). *Practical management science* (2nd edn.). Pacific Grove, CA: Duxbury.

Wright, J. R., & Houck, M. H. (1995). Water resources planning and managementm, Chap. 37. In W. F. Chen (Ed.), *Civil engineering handbook*. Boca Raton: CRC.

Wurbs, R. A. (1996). *Modeling and analysis of reservoir system operations*. Prentice Hall: Upper Saddle River, NJ.

Exercises

Engineering economics:

4.1 Consider two alternative water resource projects, A and B. Project A will cost $2,533,000 and will return $1,000,000 at the end of 5 years and $4,000,000 at the end of 10 years. Project B will cost $4,000,000 and will return $2,000,000 at the end of 5 and 15 years, and another $3,000,000 at the end of 10 years. Project A has a life of 10 years, and B has a life of 15 years. Assuming an interest rate of 0.1 (10%) per year:

(a) What is the present value of each project?

(b) What is each project's annual net benefit?

(c) Would the preferred project differ if the interest rates were 0.05?

(d) Assuming that each of these projects would be replaced with a similar project having the same time stream of costs and returns, show that by extending each series of projects to a common terminal year (e.g., 30 years), the annual net benefits of each series of projects will be same as found in part (b).

4.2 Show that $A \sum_{t=1}^{T} (1+r)^{-t} = \frac{(1+r)^T - 1}{r(1+r)^T} A$, the present value of a series of equal payments, A, at the end of each year for T years. What is the impact of an increasing interest rate over time on the present value?

4.3 (a) Show that if compounding occurs at the end of m equal length periods within a year in which the nominal interest rate is r, then the *effective* annual interest rate, r', is equal to

$$r' = \left(1 + \frac{r}{m}\right)^m - 1$$

(b) Show that when compounding is continuous (i.e., when the number of periods $m \to \infty$), the compound interest factor required to convert a present value to a future value in year T is e^{rT}. [*Hint*: Use the fact that $\lim_{k \to \infty} (1 + 1/k)^k = e$, the base of natural logarithms.]

4.4 The term "capitalized cost" refers to the present value PV of an infinite series of end-of-year equal payments, A. Assuming an interest rate of r, show that as the terminal period $T \to \infty$, PV = A/r.

4.5 The *internal rate of return* of any project or plan is the interest rate that equals the present value of all receipts or income with the present value of all costs. Show that the internal rate of return of projects A and B in Exercise 4.1 are approximately 8 and 6%, respectively. These are the interest rates r, for each project, that essentially satisfy the equation

$$\sum_{t=0}^{T} (R_t - C_t)(1+r)^{-t} = 0$$

4.6 In Exercise 4.1, the maximum annual benefits were used as an economic criterion for plan selection. The maximum benefit–cost ratio, or annual benefits divided by annual costs, is another criterion. Benefit–cost ratios should be no less than one if the annual benefits are to exceed the annual costs. Consider two projects, I and II:

	Project	
	I	II
Annual benefits	20	2
Annual costs	18	1.5
Annual net benefits	2	0.5
Benefit–cost ratio	1.11	1.3

What additional information is needed before one can determine which project is the most economical project?

4.7 Bonds are often sold to raise money for water resources project investments. Each bond is a promise to pay a specified amount of interest, usually semiannually, and to pay the face value of the bond at some specified future date. The selling price of a bond may differ from its face value. Since the interest payments are specified in advance, the current market interest rates dictate the purchase price of the bond.

Consider a bond having a face value of $10,000, paying $500 annually for 10 years. The bond or "coupon" interest rate based on its face value is 500/10,000, or 5%. If the bond is purchased for $10,000, the actual interest rate paid to the owner will equal the bond or "coupon" rate. But suppose that one can invest money in similar quality (equal risk) bonds or notes and receive 10% interest. As long as this is possible, the $10,000, 5% bond will not sell in a competitive market. In order to sell it, its purchase price has to be such that the actual interest rate paid to the owner will be 10%. In this case, show that the purchase price will be $6927.

The interest paid by the some bonds, especially municipal bonds, may be exempt from state and federal income taxes. If an investor is in the 30% income tax bracket, for example, a 5% municipal tax-exempt bond is equivalent to about a 7% taxable bond. This tax exemption helps reduce local taxes needed to pay the interest on municipal bonds, as well as providing attractive investment opportunities to individuals in high tax brackets.
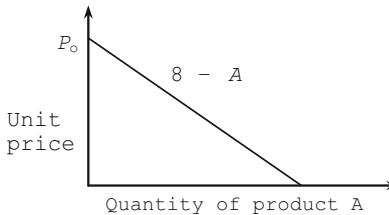
Lagrange Multipliers

4.8 What is the meaning of the Lagrange multiplier associated with the following model?

$$\text{Maximize} \quad Benefit(X) - \text{Cost}(X)$$
$$\text{Subject to:} \qquad X \le 23$$

4.9 Assume water can be allocated to three users. The allocation, x_j, to each use j provides the following returns: $R(x_1) = (12x_1 - x_1^2)$, $R(x_2) = (8x_2 - x_2^2)$ and $R(x_3) = (18x_3 - 3x_3^2)$. Assume that the objective is to maximize the total return, $F(X)$, from all three allocations and that the sum of all allocations cannot exceed 10. (a) How much would each use like to have? (b) Show that at the maximum total return solution the marginal values, $\partial(R(x_j))/\partial x_j$, are each equal to the shadow price or Lagrange multiplier (dual variable) λ associated with the constraint on the amount of water available. (c) Finally, without resolving a Lagrange multiplier problem, what would the solution be if 15 units of water were available to allocate to the three users and what would be the value of the Lagrange multiplier?

4.10 In Exercise 4.9, how would the Lagrange multiplier procedure differ if the objective function, $F(X)$, were to be minimized?

4.11 Assume that the objective was to minimize the sum of squared deviations of the actual allocations x_j from some desired or known target allocations T_j. Given a supply of water Q less than the sum of all target allocations T_j, structure a planning model and its corresponding Lagrangian. Will a global minimum be obtained from solving the partial differential equations derived from the Lagrangian? Why?

4.12 Using Lagrange multipliers, prove that the least-cost design of a cylindrical storage tank of any volume $V > 0$ has one-third of its cost in its base and top and two-thirds of its cost in its side, regardless of the cost per unit area of its base or side. (It is these types of rules that end up in handbooks in engineering design.)

4.13 An industrial firm makes two products, A and B. These products require water and
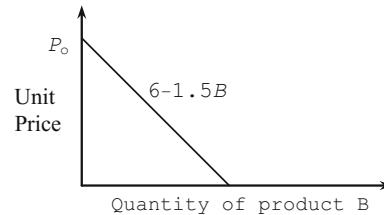
other resources. Water is the scarce resource—they have plenty of other needed resources. The products they make are unique, and hence they can set the unit price of each product at any value they want to. However experience tells them that the higher the unit price for a product, the less amount of that product they will sell. The relationship between unit price and quantity that can be sold is given by the following two demand functions. Assume for simplicity that the unit price for product A is $(8 - A)$ and for product B is $(6 - 1.5B)$.

(a) What are the amounts of A and B, and their unit prices, that maximize the total revenue obtained?

(b) Suppose the total amount of A and B could not exceed some amount T^{max}. What are the amounts of A and B, and their unit prices, that maximize total revenue, if

 (i) $T^{max} = 10$
 (ii) $T^{max} = 5$

Water is needed to make each unit of A and B. The production functions relating the amount of water X_A needed to make A, and the amount of water X_B needed to make B, are $A = 0.5X_A$, and $B = 0.25X_B$, respectively.

(c) Find the amounts of A and B and their unit prices that maximize total revenue assuming the total amount of water available is 10 units.

(d) What is the value of the dual variable, or shadow price, associated with the 10 units of available water?

Dynamic programming

4.14 Solve for the optimal integer allocations x_1, x_2, and x_3 for the problem defined by Exercise 4.9 assuming the total available water is 3 and 4. Also solve for the optimal allocation policy if the total water available is 7 and each x_j must not exceed 4.

4.15 Consider a three-season reservoir operation problem. The inflows are 10, 50 and 20 in seasons 1, 2, and 3, respectively.

Find the operating policy that minimizes the sum of total squared deviations from a constant storage target of 20 and a constant release target of 25 in each of the three seasons. Develop a discrete dynamic programming model that considers only 4 discrete storage values: 0, 10, 20 and 30. Assume the releases cannot be less than 10 or greater than 40. Show how the model's recursive equations change depending on whether the decisions are the releases or the final storage volumes. Verify the optimal operating policy is the same regardless of whether the decision variables are the releases or the final storage volumes in each period. Which model do you think is easier to solve? How would each model change if more importance were given to the desired releases than to the desired storage volumes?

4.16 Show that the constraint limiting a reservoir release, r_t, to be no greater than the

initial storage volume, s_t, plus inflow, i_t, is redundant to the continuity equation $s_t + i_t - r_t = s_{t+1}$.

4.17 Develop a general recursive equation for a forward-moving dynamic programming solution procedure for a single reservoir-operating problem. Define all variables and functions used. Why is this not a very useful approach to finding a reservoir-operating policy?

4.18 The following table provides estimates for the recent values of the costs of additional wastewater treatment plant capacity needed at the end of each 5-year period for the next 20 years. Find the capacity expansion schedule that minimizes the present values of the total future costs. If there is more than one least-cost solution, indicate which one you think is better, and why?

		Discounted cost of additional capacity					Total required capacity at end of period
		Units of additional capacity					
Period years		2	4	6	8	10	
1	1–5	12	15	18	23	26	2
2	6–10	8	11	13	15		6
3	11–15	6	8				8
4	16–20	4					10

The cost in each period t must be paid at the beginning of the period. What was the discount factor used to convert the costs at the beginning of each period t to present value costs shown above? In other words how would a cost at the beginning of period t be discounted to the beginning of period 1, given an annual interest rate of r? (Only the algebraic expression of the discount factor is asked, not the numerical value of r.)

4.19 Consider a wastewater treatment plant in which it is possible to include five different treatment processes in series. These treatment processes must together remove at least 90% of the 100 units of influent waste. Assuming the R_i is the amount of waste removed by process i, the following conditions must hold:

$$20 \leq R_1 \leq 30$$
$$0 \leq R_2 \leq 30$$
$$0 \leq R_3 \leq 10$$
$$0 \leq R_4 \leq 20$$
$$0 \leq R_5 \leq 30$$

(a) Write the constrained optimization-planning model for finding the least-cost combination of the removals R_i that together will remove 90% of the influent waste. The cost of the various discrete sizes of each unit process i depend upon the waste entering the process i as well as the amount of waste removed, as indicated in the table below.

Process i		1	2	3	4	5
Influent, I_i	Removal, R_i	Annual cost $= C_i(I_i, R_i)$				
100	20	5				
100	30	10				
80	10			3	3	1
80	20			9		2
80	30			13		
70	10			4	5	2
70	20			10		3
70	30			15		
60	10				6	2	3
60	20					4	6
60	30						9
50	10				7	3	4
50	20					5	8
50	30						10

(continued)

Process i		1	2	3	4	5
Influent, I_i	Removal, R_i	Annual cost = $C_i(I_i, R_i)$				
40	10			8	5	5
40	20				7	12
40	30					18
30	10				8	8
30	20				10	12
20	10					8

(b) Draw the dynamic programming network and solve this problem by dynamic programming. Indicate on the network the calculations required to find the least-cost path from state 100 at stage 1 to state 10 at stage 6 using both forward- and backward-moving dynamic programming solution procedures.

(c) Could the following conditions be included in the original dynamic programming model and still be solved without requiring R_4 to be 0 in the first case and R_3 to be 0 in the second case?

(i) $R_4 = 0$ if $R_3 = 0$, or
(ii) $R_3 = 0$ if $R_2 \leq 20$.

4.20 The city of Eutro Falls is under a court order to reduce the amount of phosphorus that which it discharges in its sewage to Lake Algae. The city presently has three wastewater treatment plants. Each plant i currently discharges P_i kg/day of phosphorus into the lake. Some or all plants must reduce their discharges so that the total for the three plants does not exceed P kg/day. Let X_i be the fraction or percent of the phosphorus removed by additional treatment at plant i, and the $C_i(X_i)$ the cost of such treatment ($/year) at each plant i.

(a) Structure a planning model to determine the least-cost (i.e., a cost effective) treatment plant for the city.
(b) Restructure the model for the solution by dynamic programming. Define the stages, states, decision variables, and the recursive equation for each stage.

(c) Now assume $P_1 = 20$; $P_2 = 15$; $P_3 = 25$; and $P = 20$. Make up some cost data and check the model if it works.

4.21 Find (draw) a *rule curve* for operating a single reservoir that maximizes the sum of the benefits for flood control, recreation, water supply and hydropower. Assume the average inflows in four seasons of a year are 40, 80 60, 20, and the active reservoir capacity is 100. For an average storage S and for a release of R in a season, the hydropower benefits are 2 times the square root of the product of S and R, $2(SR)^{0.5}$, and the water supply benefits are $3R^{0.7}$ in each season. The recreation benefits are $40 - (70 - S)^2$ in the third season. The flood control benefits are $20 - (40 - S)^2$ in the second season. Specify the dynamic programming recursion equations you are using to solve the problem.

4.22 How would the model defined in Exercise 4.21 change if there were a water user upstream of this reservoir and you were to find the best water-allocation policy for that user, assuming known benefits associated with these allocations that are to be included in the overall maximum benefits objective function?

4.23 Suppose there are four water users along a river who benefit from receiving water from the river. Each has a water target, i.e., each expects and plans for a specified amount. These known water targets are $W(1)$, $W(2)$, $W(3)$, and $W(4)$ for the four users, respectively. Show how dynamic programming can be used to find two allocation policies. One is to be based on minimizing the maximum deficit deviation from any target allocation. The other is to be based on minimizing the maximum percentage deficit from any target allocation.

Gradient "Hill-climbing" methods

4.24 Solve Exercise 4.13(b) using hill-climbing techniques and assuming discrete integer values and $T^{max} = 5$. For example, which product would you produce if you could make only 1 unit of either A or B? If you could make another unit of A or B, which would you make? Continue this process up to 5 units of products A and/or B.

4.25 Under what conditions will hill-climbing methods for maximization or minimization not work?
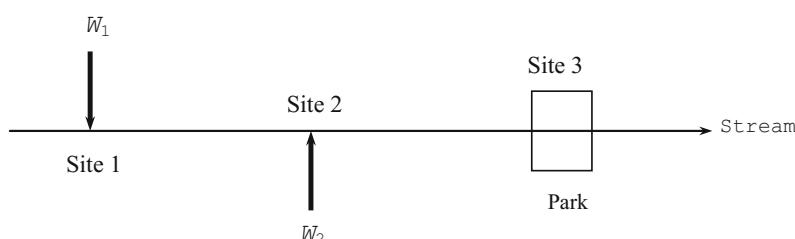
Linear and nonlinear programming

4.26 Consider the industrial firm that makes two products A and B as described in Exercise 4.13(b). Using *Lingo* (or any other program you wish):

(a) Find the amounts of A and B and their unit prices that maximize total revenue assuming the total amount of water available is 10 units.

(b) What is the value of the dual variable, or shadow price, associated with the 10 units of available water?

(c) Suppose the demand functions are not really certain. How sensitive are the allocations of water to changes in the parameter values of those functions? How sensitive are the allocations to the parameter values in the production functions?

4.27 Assume that there are m industries or municipalities that discharge their wastes into a river. Denote the discharge sites by the subscript i and let W_i be the kg of waste discharged into the river each day at those sites i. To improve the river water quality downstream, wastewater treatment plants may be required at each site i. Let x_i be the fraction of waste removed by treatment at each site i. Develop a model for estimating how much waste removal is required at each site to maintain acceptable water quality in the river at a minimum total cost. Use the following additional notation:

a_{ij} decrease in quality at site j per unit of waste discharged at site i

q_j quality at site j that would result if all controlled upstream discharges were eliminated (i.e., $W_1 = W_2 = 0$)

Q_j minimum acceptable quality at site j

C_i cost per unit (fraction) of waste removed at site i.

4.28 Assume that there are two sites along a stream, $i = 1, 2$, at which waste (BOD) is discharged. Currently, without any wastewater treatment, the quality (DO), q_2 and q_3, at each of sites 2 and 3 is less than the minimum desired, Q_2 and Q_3, respectively. For each unit of waste removed at site i upstream of site j, the quality improves by A_{ij}. How much treatment is required at sites 1 and 2 that meets the standards at a minimum total cost?

Following are the necessary data:
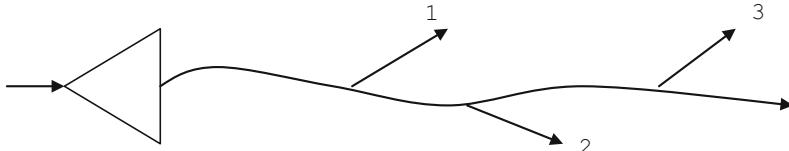
C_i cost per unit fraction of waste treatment at site i (both C_1 and C_2 are unknown but for the same amount of treatment, whatever that amount, $C_1 > C_2$)

R_i decision variables, unknown waste removal fractions at sites $i = 1, 2$

$$A_{12} = 1/20 \quad W_1 = 100 \quad Q_2 = 6$$
$$A_{13} = 1/40 \quad W_2 = 75 \quad Q_3 = 4$$
$$A_{23} = 1/30 \quad q_2 = 3 \quad q_3 = 1$$

4.29   Define a linear programming model for finding the tradeoff between active storage capacity and the maximum percentage deviation from a known target storage volume and a known target release in each period. How could the solution of the model be used to define a reservoir policy?

4.30   Consider the possibility of building a reservoir upstream of three demand sites along a river.

guaranteed in each season) and the reservoir capacity.

(b) Find the tradeoff between the yield and the maximum total net benefits that can be obtained from allocating that yield among the three users.

(c) Find the tradeoff between the reservoir capacity and the total net benefits one can obtain from allocating the total releases, not just the reliable yield, to the downstream users.

(d) Assuming a reservoir capacity of 7, and dividing the release into integer increments of 2 (i.e., 2, 4, 6 and 8), using linear programming, find the optimal operating policy. Assume the maximum release cannot exceed 8, and the minimum release cannot be less than 2. How does this solution differ from that obtained using dynamic programming?

(e) If you were maximizing the total net benefit obtained from the three users and if the water available to allocate to

The net benefits derived from each use depend on the reliable amounts of water allocated to each use. Letting x_{it} be the allocation to use i in period t, the net benefits for each period t equal

1. $6x_{1t} - x_{1t}^2$
2. $7x_{2t} - 1.5x_{2t}^2$
3. $8x_{3t} - 0.5x_{3t}^2$

Assume the average inflows to the reservoir in each of four seasons of the year equal 10, 2, 8, 12.

(a) Find the tradeoff between the yield (the expected release that can be

the three users were 15 in a particular time period, what would be the value of the Lagrange multiplier or dual variable associated with the constraint that you cannot allocate more than 15 to the three uses?

(f) There is the possibility of obtaining recreational benefits in seasons 2 and 3 from reservoir storage. No recreational benefits can occur in seasons 1 and 4. To obtain these benefits facilities must be built, and the question is at what elevation (storage volume) should they be built. This is called the recreational storage volume target.

Recreational benefits in each recreation season equal 8 per unit of storage target if the actual storage equals the storage target. If the actual storage is less than the target the losses are 12 per unit deficit—the difference between the target and actual storage volumes. If the actual storage volume is greater than the target volume the losses are 4 per unit excess. What is the reservoir capacity and recreation storage target that maximizes the annual total net benefits obtained from downstream allocations and recreation in the reservoir less the annual cost of the reservoir, $3K^{1.5}$, where K is the reservoir capacity?

(g) In (f) above, suppose the allocation benefits and net recreation benefits were given weights indicating their relative importance. What happens to the relationship between capacity K and recreation target as the total allocation benefits are given a greater weight in comparison to recreation net benefits?

4.31 Using the network representation of the wastewater treatment plant design problem defined in Exercise 4.19, write a linear programming model for defining the least-cost sequence of unit treatment process (i.e., the least-cost path through the network). [*Hint*: Let each decision variable x_{ij} indicate whether or not the link between nodes (or states) i and j connecting two successive stages is on the least-cost or optimal path. The constraints for each node must ensure that what enters the node must also leave the node.]

4.32 Two types of crops can be grown in particular irrigation area each year. Each unit quantity of crop A can be sold for a price P_A and requires W_A units of water, L_A units of land, F_A units of fertilizer, and H_A units of labor. Similarly, crop B can be sold at a unit price of P_B and requires W_B, L_B, F_B and H_B units of water, land, fertilizer, and

labor, respectively, per unit of crop. Assume that the available quantities of water, land, fertilizer, and labor are known, and equal W, L, F, and H, respectively.

(a) Structure a linear programming model for estimating the quantities of each of the two crops that should be produced in order to maximize total income.

(b) Solve the problem graphically, using the following data:

Resource	Requirements per unit of		Maximum available resources
	Crop A	Crop B	
Water	2	3	60
Land	5	2	80
Fertilizer	3	2	60
Labor	1	2	40
Unit price	30	25	

(c) Define the meaning of the dual variables, and their values, associated with each constraint.

(d) Write the dual model of this problem and interpret its objective and constraints.

(e) Solve the primal and dual models using an existing computer program, and indicate the meaning of all output data.

(f) Assume that one could purchase additional water, land, fertilizer, and labor with capital that could be borrowed from a bank at an annual interest rate r. How would this opportunity alter the linear programming model? The objective continues to be a maximization of net income. Assume there is a maximum limit on the amount of money that can be borrowed from the bank.

(g) Assume that the unit price P_j of crop j is a decreasing linear function $(P_j^o - b_j x_j)$ of the quantity, x_j, produced. How could the linear model be

restructured also as to identify not only how much of each crop to produce, but also the unit price at which each crop should be sold in order to maximize total income?

4.33 Using linear programming model, derive an annual storage-yield function for a reservoir at a site having the following record of annual flows:
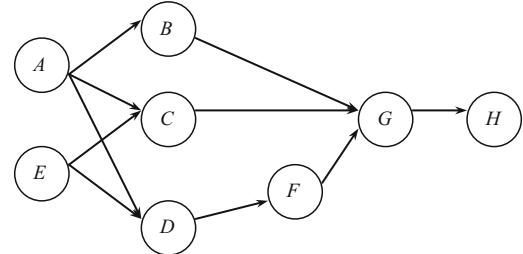
Year y	Flow Q_y	Year y	Flow Q_y
1	5	9	3
2	7	10	6
3	8	11	8
4	4	12	9
5	3	13	3
6	3	14	4
7	2	15	9
8	1		

(a) Find the values of the storage capacity required for yields of 2, 3, 3.5, 4, 4.5, and 5.
(b) Develop a flow chart defining a procedure for finding the yields for various increasing values of K.

4.34 Water resources planning usually involves a set of separate tasks. Let the index i denote each task, and H_i the set of tasks that immediately precede task i. The duration of each task i is estimated to be d_i.

(a) Develop a linear programming model to identify the starting times of tasks that minimizes the time, T, required to complete the total planning project.
(b) Apply the general model to the following planning project:

Task A: Determine planning objectives and stakeholder interests. Duration: 4 months.

Task B: Determine structural and nonstructural alternatives that will influence objectives. Duration: 1 month.

Task C: Develop an optimization model for preliminary screening of alternatives and for estimating tradeoffs among objectives. Duration: 1 month.

Task D: Identify data requirements and collect data. Duration: 2 months.

Task E: Develop a data management system for the project. Duration: 3 months.

Task F: Develop an interactive shared vision simulation model with the stakeholdes. Duration: 2 Months.

Task G: Work with stakeholders in an effort to come to a consensus (a shared vision) of the best plan. Duration: 4 months.

Task H: Prepare, present and submit a report. Duration: 2 months.
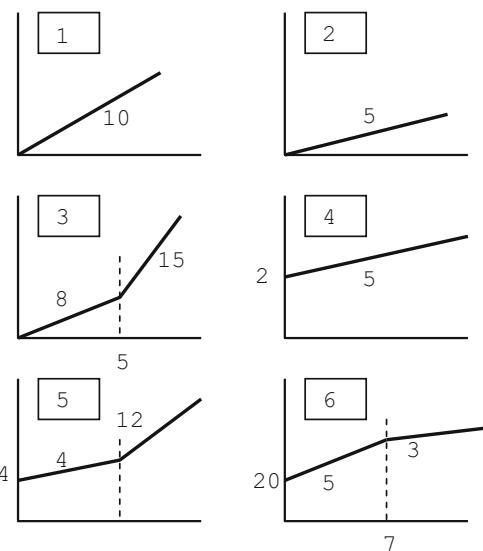
4.35 In Exercise 4.34 suppose the project is penalized if its completion time exceeds a target T. The difference between 14 months and T months is Δ, and the penalty is $P(\Delta)$. You could reduce the time it takes to complete task E by one month at a cost of $200, and by two months at a cost of $500. Similarly, suppose the cost of task A could be reduced by a month at a cost of $600 and two months at a cost of $1400. Construct a model to find the most economical project completion time. Next modify the linear programming model to find the minimum total added cost if the total project time is to be reduced by 1 or 2 months. What is that added cost and for which tasks?

4.36 Solve the reservoir operation problem described in Exercise 4.15 using linear programming. If the reservoir capacity is unknown, show how a cost function (that includes fixed costs and economies of scale) for the reservoir capacity could be included in the linear programming model.

4.37 An upstream reservoir could be built to serve two downstream users. Each user has a constant water demand target. The first user's target is 30; the second user's target is 50. These targets apply to each of 6 within-year seasons. Find the tradeoff between the required reservoir capacity and maximum deficit to any user at any time, for an average year. The average flows into the reservoir in each of the six successive seasons are: 40, 80, 100, 130, 70, 50.

4.38 Two groundwater well fields can be used to meet the water demands of a single user. The maximum capacity of the *A* well field is 15 units of water per period, and the maximum capacity of the *B* well field is 10 units of water per period. The annual cost of building and operating each well field, each period, is a function of the amount of water pumped and transported from that well field. Three sets of cost functions are shown below: Construct a LP model and use it to define and then plot the total least-cost function and the associated individual well field capacities required to meet demands from 0 to 25, assuming cost functions 1 and 2 apply to well fields *A* and *B,* respectively. Next define another least-cost function and associated capacities assuming cost functions 3 and 4 apply to *A* and *B,* respectively. Finally define a least-cost function and associated capacities assuming well field cost functions 5 and 6 apply. You can check your model results just using common sense—the least-cost functions should be obvious, even without using optimization.

4.39 Referring to Exercise 4.38 above, assume cost functions 5 and 6 represent the cost of adding additional capacity to well fields *A* and *B,* respectively, in any of the next five 5-year construction periods, i.e., in the next 25 years. Identify and plot the least-cost capacity expansion schedule (one that minimizes the total present value of current and future expansions), assuming demands of 5, 10, 15, 20 and 25 are to be met at the end of years 5, 10, 15, 20 and 25, respectively. Costs, including fixed costs, of capacity expansion in each construction period have to be paid at the beginning of the construction period. Determine the sensitivity of your solution to the interest rate used to compute present value.

4.40 Consider a crop production problem involving three types of crops. How many hectares of each crop should be planted to maximize total income?

Resources	Max limits	Resource requirements			
		Crops:	Corn	Wheat	Oats	
Water	1000/week		3.0	1.0	1.5	units/week/ha
Labor	300/week		0.8	0.2	0.3	person h/week/ha
Land	625 ha					
Yield $/ha			400	200	250	

Show a graph that identifies the tradeoffs among crops that can be made without reducing the total income.

4.41 Releases from a reservoir are used for water supply or for hydropower. The benefit per unit of water allocated to hydropower is BH and the benefit per unit of water allocated to water supply is BW. For any given release the difference between the allocations to the two uses cannot exceed 50% of the total amount of water available. Show graphically how to determine the most profitable allocation of the water for some assumed values of BH and BW. From the graph identify which constraints are binding and what their "dual prices" mean (in words).

4.42 Suppose there are four water users along a river who benefit from receiving water. Each has a known water target, i.e., each expects and plans for a specified amount. These known water targets are W_1, W_2, W_3, and W_4 for the four users, respectively. Find two allocation policies. One is to be based on minimizing the maximum deficit deviation from any target allocation. The other is to be based on minimizing the maximum percentage deficit from any target allocation.

Deficit allocations are allocations that are less than the target allocation. For example if a target allocation is 30 and the actual allocation is 20, the deficit is 10. Water in excess of the targets can remain in the river. The policies are to indicate what the allocations should be for any particular river flow Q. The policies can be expressed on a graph showing the amount of Q on the horizontal axis, and each user's allocation on the vertical axis.

Create the two optimization models that can be used to find the two policies and indicate how they would be used to define the policies. What are the unknown variables and what are the known variables? Specify the model in words as well as mathematically.

4.43 In Indonesia there exists a wet season followed by a dry season each year. In one area of Indonesia all farmers within an irrigation district plant and grow rice during the wet season. This crop brings the farmer the largest income per hectare; thus they would all prefer to continue growing rice during the dry season. However, there is insufficient water during the dry season to irrigate all 5000 ha of available irrigable land for rice production. Assume an available irrigation water supply of 32×10^6 m^3 at the beginning of each dry season, and a minimum requirement of 7000 m^3/ha for rice and 1800 m^3/ha for the second crop.

(a) What proportion of the 5000 ha should the irrigation district manager allocate for rice during the dry season each year, provided that all available hectares must be given sufficient water for rice or the second crop?

(b) Suppose that crop production functions are available for the two crops, indicating the increase in yield per hectare per m^3 of additional water, up to 10, 000 m^3/ha for the second crop. Develop a model in which the water allocation per hectare, as well as the hectares allocated to each crop, is to be determined, assuming a specified price or return per unit of yield of each crop. Under what conditions would the solution of this model be the same as in part (a)?

4.44 Along the Nile River in Egypt, irrigation farming is practiced for the production of cotton, maize, rice, sorghum, full and short

berseem for animal production, wheat, barley, horsebeans, and winter and summer tomatoes. Cattle and buffalo are also produced, and together with the crops that require labor, water. Fertilizer, and land area (feddans). Farm types or management practices are fairly uniform, and hence in any analysis of irrigation policies in this region this distinction need not be made. Given the accompanying data develop a model for determining the tons of crops and numbers of animals to be grown that will maximize (a) net economic benefits based on Egyptian prices, and (b) net economic benefits based on international prices. Identify all variables used in the model. Known parameters:

C_i miscellaneous cost of land preparation per feddan

P_i^E Egyptian price per 1000 tons of crop i

P_i^I international price per 1000 tons of crop i

v value of meat and dairy production per animal

g annual labor cost per worker

f^P cost of P fertilizer per ton

f^N cost of N fertilizer per ton

Y_i yield of crop i, tons/feddan

α feddans serviced per animal

β tons straw equivalent per ton of berseem carryover from winter to summer

r^w berseem requirements per animal in winter

s^{wh} straw yield from wheat, tons per feddan

s^{ba} straw yield from barley, tons per feddan

r^s straw requirements per animal in summer

μ_i^N N fertilizer required per feddan of crop i

μ_i^P P fertilizer required per feddan of crop i

l_{im} labor requirements per feddan in month m, man-days

w_{im} water requirements per feddan in month m, 1000 m^3

h_{im} land requirements per month, fraction (1 = full month)

Required Constraints (assume known resource limitations for labor, water, and land):

(a) Summer and winter fodder (berseem) requirements for the animals.

(b) Monthly labor limitations.

(c) Monthly water limitations.

(d) Land availability each month.

(e) Minimum number of animals required for cultivation.

(f) Upper bounds on summer and winter tomatoes (assume these are known).

(g) Lower bounds on cotton areas (assume this is known).

Other possible constraints:

(a) Crop balances.

(b) Fertilizer balances.

(c) Labor balance.

(d) Land balance.

4.45   In Algeria there are two distinct cropping intensities, depending upon the availability of water. Consider a single crop that can be grown under intensive rotation or extensive rotation on a total of A hectares. Assume that the annual water requirements for the intensive rotation policy are 16,000 m^3 per ha, and for the extensive rotation policy they 4000 m^3 per ha. The annual net production returns are 4000 and 2000 dinars, respectively. If the total water available is 320,000 m^3, show that as the available land area A increases, the rotation policy that maximizes total net income changes from one that is totally intensive to one that is increasingly extensive.

Would the same conclusion hold if instead of fixed net incomes of 4000 and 2000 dinars per hectares of intensive and extensive rotation, the net income depended on the quantity of crop produced? Assuming that intensive rotation produces twice as much produced by extensive rotation, and that the net income per unit of crop Y is defined by the simple linear function $5 - 0.05\ Y$, develop and solve a linear programming model to determine the optimal rotation policies if A equals 20, 50, and 80. Need this net income or

price function be linear to be included in a linear programming model?

4.46 Current stream quality is below desired minimum levels throughout the stream in spite of treatment at each of the treatment plant and discharge sites shown below. Currently effluent standards are not being

the amount of wastewater to treat at each site and then release to the river. Any wastewater at any site that is not undergoing additional treatment can be piped to other sites. Identify other issues that could affect the eventual decision.

○ Wastewater Treatment Plant

▪▪▪▪▪ Possible pipeline for wastewater effluent.

met, and minimum desired streamflow concentrations can be met by meeting effluent standards. All current wastewater discharges must undergo additional treatment. The issue is where additional treatment is to occur and how much.

Develop a model to identify cost-effective options for meeting effluent standards where ever wastewater is discharged into the stream. The decisions variables include

Assume known current wastewater flows at site $i = q_i$.

Additional treatment to meet effluent standards cost $= a_i + b_i(D_i)^{c_i}$ where D_i is the total wastewater flow undergoing additional treatment at site i and $c_i < 1$.

Pipeline and pumping for each pipeline segment costs approximately $\alpha_{ij} + \beta(q_{ij})^{\gamma}$.

where q_{ij} is pipeline flow between adjacent sites i and j and $\gamma < 1$.

Clearly, all model outputs depend on model inputs. The optimization and simulation models discussed in the previous chapters are no exception. This chapter introduces some alternative modeling approaches that depend on observed data. These approaches include artificial neural networks and various evolutionary models. The chapter ends with some qualitative modeling. These data-driven models can serve as substitutes for more process-based models in applications where computational speed is critical or where the underlying relationships are poorly understood or too complex to be easily incorporated into calculus-based, linear, nonlinear, or dynamic programming models. Evolutionary algorithms involve random searches based on evolutionary or biological processes for finding the values of parameters and decision variables that best satisfy system performance criteria. Evolutionary algorithms are popular methods for analyzing systems that require complex simulation models to determine values of performance measures. Qualitative modeling approaches are useful when performance measures are expressed qualitatively, such as "I want a reliable supply of clean water at a reasonable cost," where there can be disagreements among different stakeholders and decision makers with respect to specifying just how reliable, how clean, and how affordable.

5.1 Introduction

Most 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. Such models do not describe physical processes. They attempt to convert observed inputs (e.g., rainfall and runoff, inflows to a reservoir, pollutants entering a wastewater treatment plant or effluent concentrations discharged to a river) to observed outputs (e.g., runoff, reservoir releases, pollutant

concentrations) using any set of mathematical equations or expressions that does the job. One type of such models is regression.

Regression equations, such as of the forms

$$\text{Output variable value} = a + b(\text{input variable value}) \tag{5.1}$$

$$\text{Output variable value} = a + b(\text{input variable value})^C \tag{5.2}$$
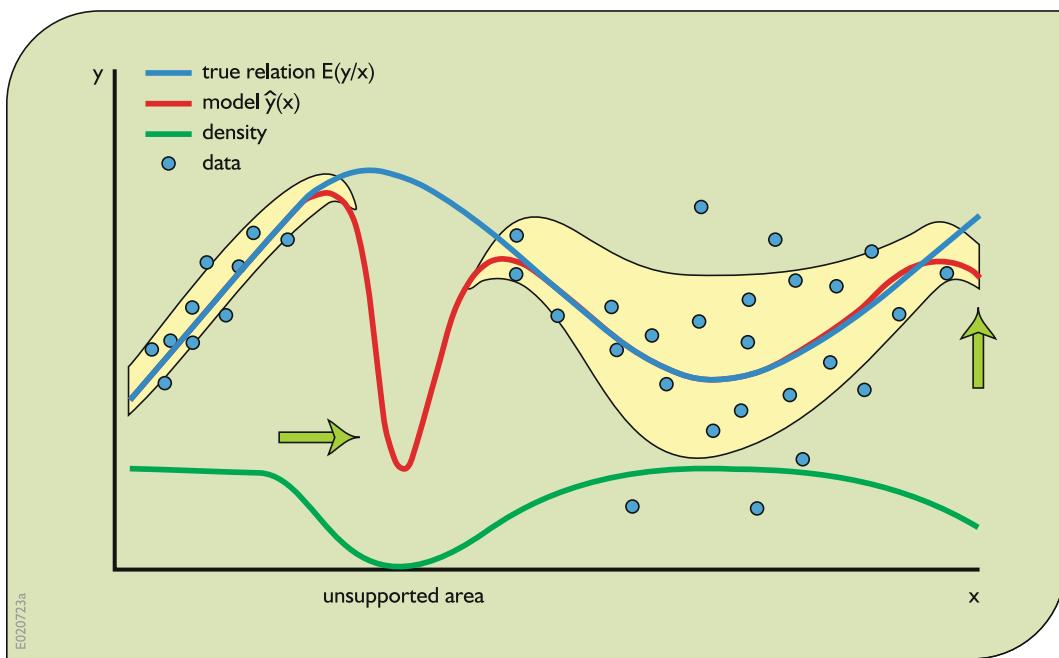
$$\begin{aligned}\text{Output variable value} \\ = a + b_1(\text{input variable}_1\text{value})^{C1} \\ + b_2(\text{input variable}_2\text{value})^{C2}\end{aligned} \tag{5.3}$$

are examples of such data-fitting or statistical models.

They depend on 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 Fig. 5.1), many have proven quite successful in representing complex physical systems.

Other examples of data-driven models are based on biological 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 their varieties used to find the values of parameters and variables that best satisfy some objective. Alternatively, an extension of regression is artificial neural networks (ANN). The development and application of black-box models like GA, GP,

Fig. 5.1 Data-fitting 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

and ANNs emulate larger, deterministic, process-oriented models. Once calibrated, their use may be advantageous if and when it is quicker to use them to obtain the information needed rather than using process-oriented models that typically take longer to solve. Process-oriented models are sometimes used to calibrate artificial neural networks, which are then used to more quickly explore and evaluate the range of solution outputs associated with varying inputs.

Examples of such situations where multiple solutions of a model must be obtained include sensitivity or uncertainty analysis, scenario evaluations, risk assessment, optimization, inverse modeling to obtain parameter values given the values of the decision variables, 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-fitting models for model emulation are given in the next several 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 sets 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 are 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 (data used for calibration) without any assumptions about the data's statistical distribution or other characteristics (Hagan et al. 1996; Hertz et al. 1991).
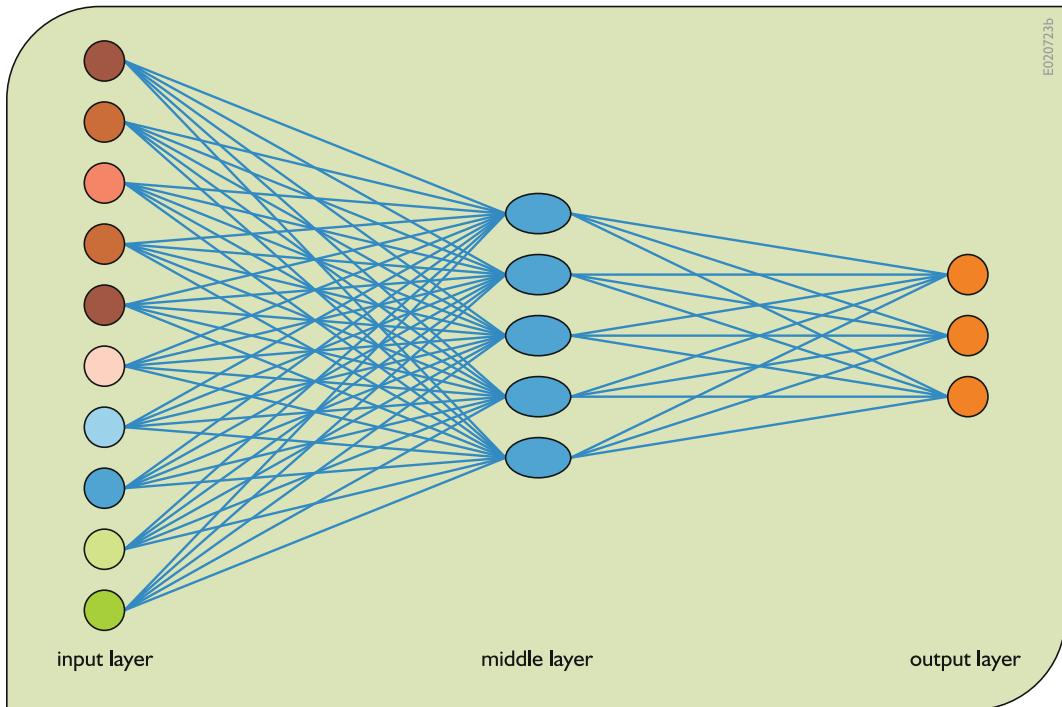
The following sections are intended to provide some background helpful to those who may be selecting one among all the available computer codes for implementing a genetic algorithm, genetic program, or artificial neural network.

5.2 Artificial Neural Networks

5.2.1 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 (e.g., see Fig. 2.1) and will continue in the future even given our continually improving electronic digital processing 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 uses ANNs representing simplified models of the brain. In reality, it is just a more complex type of regression or statistical (black-box) model.

Fig. 5.2 A typical multilayer artificial neural network showing the input layer for ten different inputs, the hidden layer (s), and the output layer having three outputs

An example of the basic structure of an ANN is shown in Fig. 5.2. 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 column(s) 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 these 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)* consist(s) 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 Fig. 5.2 has links only between nodes in immediately adjacent layers or columns and is often referred to as a multilayer perceptron (MLP) network, or a feedforward (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.

Input value to node k:

$$I_k = \sum w_j O_j \qquad (5.4)$$
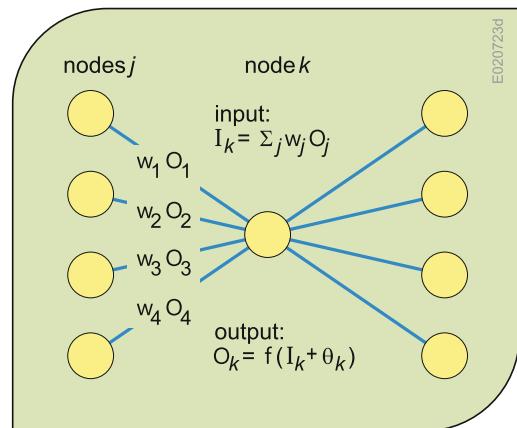
Again, the sum in Eq. 5.4 is over all nodes j providing inputs to node k.

At each node k of hidden and output layers, the input I_k is an argument to a linear or nonlinear function $f_k(I_k + \theta_k)$, which converts the input I_k to output O_k. The variable θ_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 Eq. 5.5 and graphed in Fig. 5.3.

$$O_k = 1/[1 + \exp\{-(I_k + \theta_k)\}] \qquad (5.5)$$

The process of converting inputs to outputs at each hidden layer node is illustrated in Fig. 5.4. The same process also happens at each output layer node.

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
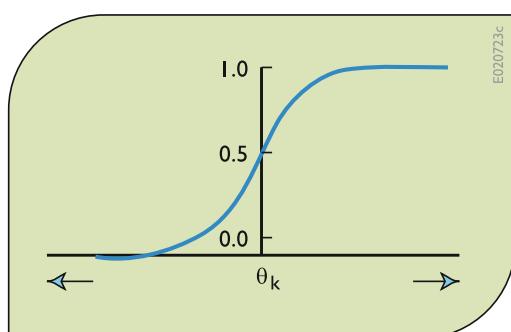
Fig. 5.4 A middle-layer node k converting input values to an output value using a nonlinear function f (such as defined by Eq. 5.5) in a multilayer ANN

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:

- *Feedforward 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 Fig. 5.2); however, this is not a requirement of feedforward 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

Fig. 5.3 The sigmoid or logistic threshold function with threshold θ_k

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 multilayer feedforward 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 30 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.

5.2.2 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 5.1.

Table 5.1 Streamflow travel times and pollutant concentrations

travel time (days)	concentration	
	upstream	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

Assume at first that the ANN structure consists of two input nodes, a hidden node, and a single output node. One of the input nodes is for the upstream concentration and the other input node is for the travel time. The single output node represents the downstream concentration expressed as a fraction of the upstream concentration. This is shown in Fig. 5.5.

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 (Eq. 5.5) is applied at the middle node and at the output node. Using two or more data sets to train or calibrate this ANN (Fig. 5.5) 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 Fig. 5.6.
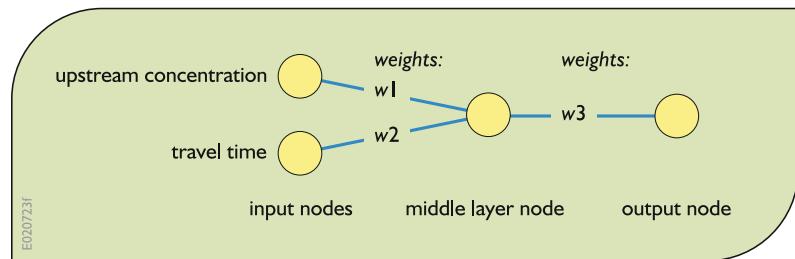
Using only half the data (six data sets) for training or calibration, the weights obtained provided a near perfect fit. The weights obtained are shown in Table 5.2.
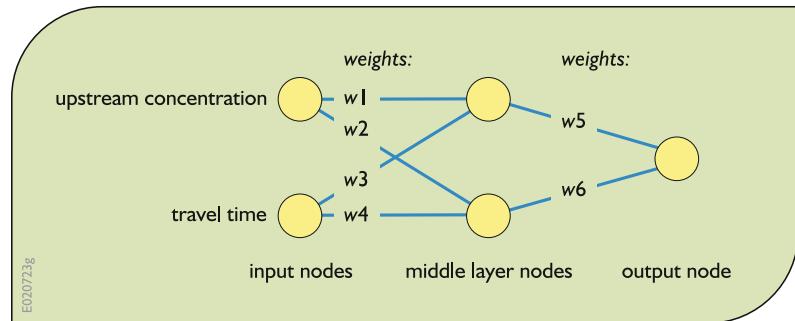
Next the remaining six data sets were applied to the network with weights set to those values shown in Table 5.2. Again the sum of absolute deviations was essentially 0. Similar results were obtained with increasing numbers of data sets.

The values of the weights in Table 5.2 indicate something water quality modelers typically assume, and that is that the fraction of the upstream pollutant concentration that reaches a downstream site is independent of the actual upstream concentration (see Chap. 4). This ANN

Fig. 5.5 Initial ANN for example problem

Fig. 5.6 Modified ANN for example problem

Table 5.2 Weights for each link of the ANN shown in Fig. 5.6 based on six data sets from Table 5.1. All bias variables (θ_k in Eq. 5.5) 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		

could have had only one input node, namely that for travel time. This conforms to the typical first-order decay function:

Fraction of pollutant concentration downstream per unit concentration upstream

$$= \exp\{-k(\text{travel time})\},$$

$$(5.6)$$

where the parameter k is the decay rate constant having units of 1/travel time (travel time units^{-1}).

5.3 Evolutionary Algorithms

Evolutionary algorithms (EA) represent a broad spectrum of heuristic approaches for simulating biological evolution in the search for improved

"fitness," i.e., the best values of decision variables and parameters based on an objective or fitness function. Evolutionary algorithms are broadly based on the repeated mutation and recombination and selection: in each generation (iteration) to define new individuals (candidate solutions). These are generated by variation, usually in a stochastic way, and then some individuals are selected for the next generation based on their relative fitness or objection function value. Over the generation sequence, individuals with increasingly better fitness values are generated (Simon 2013).
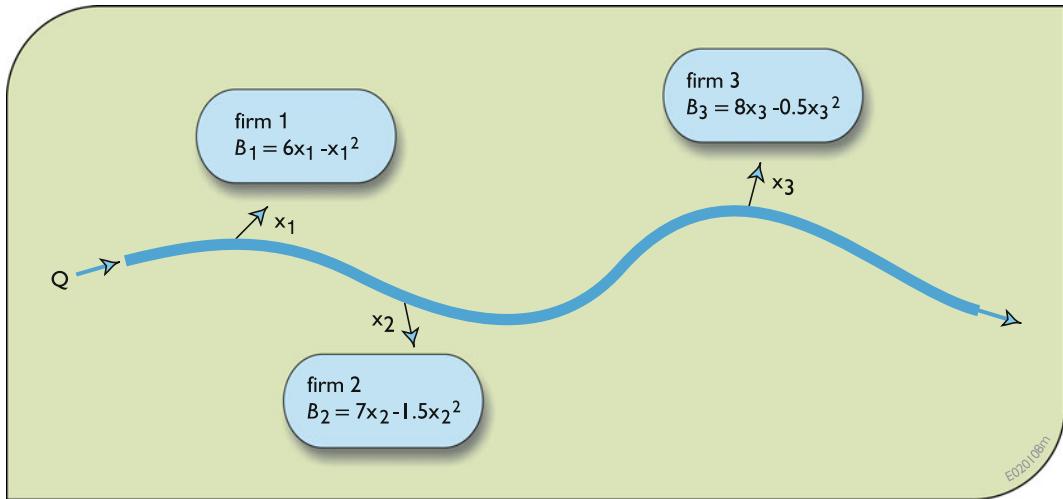
Primary examples include genetic algorithms (Holland 1975), evolutionary strategies (Rechenberg 1973; Schwefel 1981), evolutionary programming (Fogel et al. 1966), and genetic programming (Koza 1992). These methods are comprised of algorithms that operate using a population of alternative solutions or designs, each represented by a potential decision vector. They rely on randomized operators that simulate mutation and recombination to create new individuals, i.e., solutions, who then compete to survive via the selection process, which operates according to a problem-specific fitness or objective function. In some cases this function can be a complex simulation model dependent on the values of its parameters and decision variables derived from the EA. EA popularity is, at least in part, due to their potential to solve nonlinear, nonconvex, multimodal, and discrete problems for which deterministic gradient-based search techniques incur difficulty or fail completely. The growing complexity and scope of environmental and water resources applications has served to expand EAs' capabilities.

Currently, the field of biologically inspired search algorithms mostly include variations of evolutionary algorithms and swarm intelligence algorithms, e.g., ant colony optimization (ACO), particle swarm optimization (PSO), bees algorithm, bacterial foraging optimization (BFO), and so on, many of which have been used to analyze water resources planning and management problems. This is especially true for application of genetic algorithms, arguably among the most popular of the several types of EAs. EAs are flexible tools that can be applied to the solution of a wide variety of complex water resources problems. Nicklow et al. (2010) provides a comprehensive review of state-of-the-art methods and their applications in the field of water resources planning and management. EAs have been successfully applied to the study of water distribution systems, urban drainage and sewer systems, water supply and wastewater treatment, hydrologic and fluvial modeling, groundwater systems, and parameter identification, to name a few. Nicklow et al. also identify major challenges and opportunities for the future, including a call to address larger scale problems that involve uncertainty and an expanded need for collaboration among multiple stakeholders and disciplines. Evolutionary computation methods will surely continue to evolve in the future as analysts encounter increased problem complexities and uncertainty and as the societal pressure for more innovative and efficient solutions rises.

5.3.1 Genetic Algorithms

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. 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 populations of possible solutions, attempting to find a solution set that either maximizes or minimizes the value of a function of those parameters and decision variables. This function is called an 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. This process continues until no significant improvement in model output is apparent. Just how good or "fit" a particular population of parameter and decision variable values is must be evaluated using a model of the

Fig. 5.7 Water allocation to three users from a stream having a flow of Q

system that contains these parameters and decision variables. This system model is separated from the GA model. This model separation makes GA applicable for estimating the best parameter and decision variable values of a wide variety of simulation models used for planning, design, operation, and management.

Each individual solution set of a GA model 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.* Assuming 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 of 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 and not reappear in future population sets.

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 Fig. 5.7. Assume only integer solutions are to be considered. The maximum allocation, x_i, to any single user i cannot exceed 5, and the sum of all allocations cannot exceed the value of Q, say 6.

$$0 \leq x_i \leq 5 \quad \text{for} \quad i = 1, 2, \text{ and } 3. \tag{5.7}$$

$$x_1 + x_2 + x_3 \leq 6 \tag{5.8}$$

The objective is to find the values of each allocation that maximizes the total benefits, $B(X)$, while satisfying (5.7) and (5.8).

$$\text{Maximize } B(X) = (6x_1 - x_1^2) + (7x_2 - 1.5x_2^2) + (8x_3 - 0.5x_3^2) \tag{5.9}$$

A population of possible feasible solutions is generated randomly. The best size of the sample solution population—the number of solutions being considered—is 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. 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,

$$3 \mid 1\,2$$
$$1 \mid 0\,1$$

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 preset probability. For example, suppose the probability of a crossover was 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. Assume as the result of this "uniform" crossover procedure, only the middle gene in the pair of strings 312 and 101 is exchanged. This would result in 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.

Next consider mutation. Random mutation operations can apply to each gene in each string. Mutation involves changing the value of the gene being mutated. If these strings contain 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 processes have 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 of a child, the higher the probability of its being selected to be a parent in a new population.

In this example, the objective is to maximize the total benefit derived from the allocation of water, Eq. 5.9. Referring to Eq. 5.9, the string 301 has a total benefit of 16.5. The string 102 has a total benefit of 19.0. Considering just these two children, the sum of these two individual benefits is 35.5. Thus the child (string) 301 has a probability of $16.5/35.5 = 0.47$ of being selected for the new population, and the other child (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 0–0.47, 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 (more fit) 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 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 Fig. 5.8.

5.3.2   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 modeler/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.
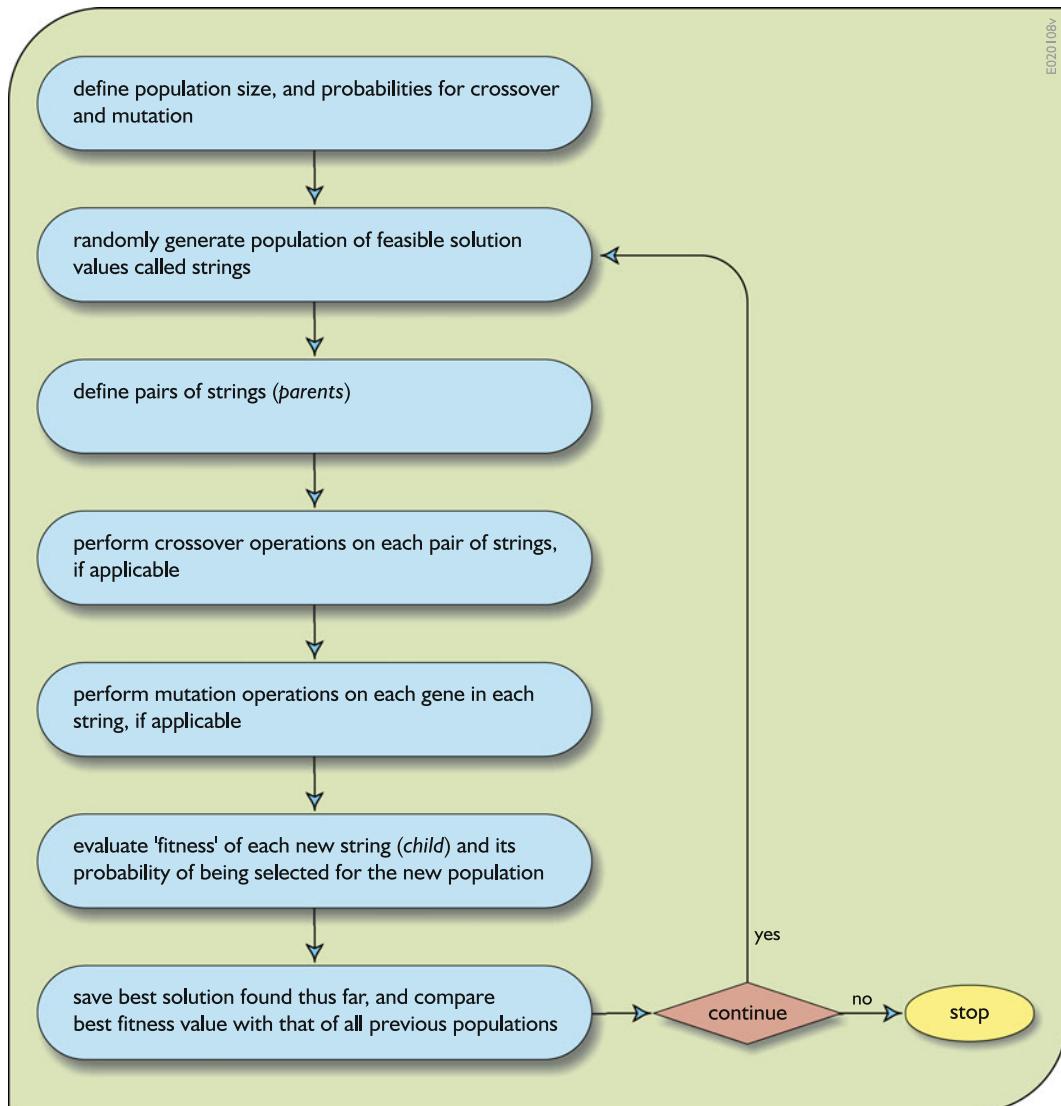
For this example, we will

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 A_1, A_2, and A_3 that sum 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 as 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 Eq. 5.9, 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 5.3 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

Fig. 5.8 Flow chart of genetic algorithm procedure

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.

Note that the above description of how genetic algorithms work permits the use of any "fitness function" for comparing alternative solutions, and for selecting preferred ones. The search procedure is independent of the particular characteristics of the water resource system being analyzed. This fitness "function" can be a complex groundwater quality model, for example, the parameter values of which are being suggested by the outcome of the GA procedure. Thus in such an application, both simulation and optimization procedures are combined and there are no restrictions on the features of either. As might

Table 5.3 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	0.91	230
	310	310	310	210	<u>13.5</u>	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	<u>26.5</u>	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			

be expected, this has opened up a wide variety of planning and management problems that now can be analyzed in the search for effective solutions.

5.3.3 Differential Evolution

Differential evolution (DE) operates through similar computational steps as employed by a standard evolutionary algorithm (EA) such as genetic algorithms. DE is an optimization technique that iteratively modifies a population of candidate solutions to make it converge to an optimum value. However, unlike traditional EAs, DE programs create new-generation population members by adding a weighted difference between two population vectors to a third vector. To illustrate, after initializing multiple candidate solutions with random values, begin an iterative process where for each candidate solution x you produce a trial vector $v = a + (b - c)/2$, where a, b, c are three distinct candidate solutions picked randomly among the population of possible solutions. Next, you randomly swap vector components between x and v to produce v'. At least one component from v must be swapped. Finally, you replace x in your population with v' only if v' is a better candidate (i.e., it improves the value your objective or fitness function). This process is repeated until no better solution can be found. No separate probability distribution need be used for generating the offspring.

Since its inception in 1995, many variants of the basic algorithm have been developed with improved performance. Books and web pages are available that present detailed reviews of the basic concepts of DE and of its major variants, as well as its application to multiobjective, constrained, large-scale, and uncertain optimization problems. Numerous computer software packages are also available for solving problems using DE. For example, see Das and Suganthan (2011), Storn and Price (1997), Price et al. (2006), and Schwefel (1995) to mention a few.

5.3.4 Covariance Matrix Adaptation Evolution Strategy

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is another stochastic, derivative-free method for numerical solution of nonlinear or nonconvex continuous optimization problems. They belong to the class of evolutionary algorithms. Pairwise dependencies between the variables are represented by a covariance matrix. The covariance matrix adaptation (CMA) method updates the covariance matrix in a way that improves the value of the fitness function. Adaptation of the covariance matrix is similar to the approximation of the inverse Hessian matrix in calculus-based optimization. In contrast to most classical methods, fewer assumptions on the nature of the underlying objective function are made. Only the ranking between candidate solutions is exploited for learning the sample distribution and neither derivatives nor even the function values themselves are required by the method (Hansen 2006; Igel et al. 2007).

Some software programs for DE are at (http://www1.icsi.berkeley.edu/~storn/code.html), for CMA-ES at (https://www.lri.fr/~hansen/cmaesintro.html) and for multiobjective EAs at (http://moeaframework.org/).

5.4 Genetic Programming

One of the challenges in computer science is to program computers to perform tasks without telling them how. In other words, how to 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 the 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 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 et al. (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 analyze 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 Eqs. 5.1, 5.2, or 5.3) will give us any insight into the actual workings of the system.

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 the basic building blocks (mathematical operations and variables) that may be used; the algorithm then tries to build the model using sequences of the specified building blocks.
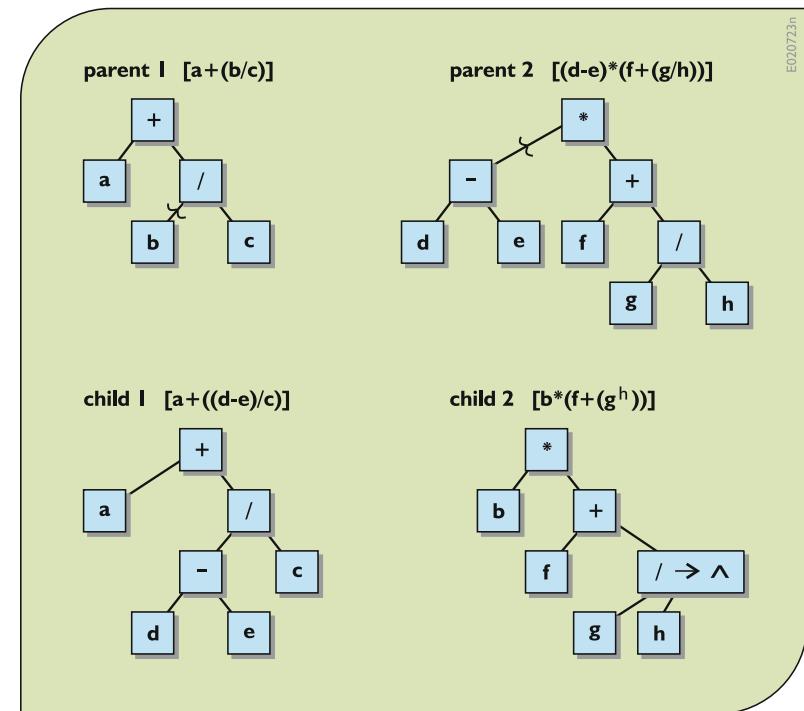
One of the successful applications of GP in automatic model building is that of symbolic regression. Here GP searches for a mathematical regression expression in symbolic form that best produces the observed output given the associated input. To perform this task GP uses a physical symbol system divided into two sets. The first set contains the symbols for independent variables as well as parameter constants as appropriate. The content of this set is based on the nature of the problem to be solved. The second set contains the basic operators used to form a function. For example, the second set can contain the arithmetic operators $(+, -, *, /)$ and perhaps others such as log, square root, sine, and cosine as well, again based on the perceived degree of complexity of the regression.

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 the parse tree for the expression $a + (b/c)$ illustrated in Fig. 5.9. The crossover operation simply exchanges a branch of one parent with a branch of the other.

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 modeling (Babovic and Abbott 1997; Drecourt 1999; Liong et al. 2000), sediment transport modeling, 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. There are also potential problems in creating a dimensionally correct model if the input data are not dimensionless. As a consequence, many applications using

Fig. 5.9 An illustration of a crossover operation and mutation operation for genetic programming

GP seem to require some guidance based on a mix of both physically based and data-driven approaches.

5.5   Qualitative Functions and Modeling

So far the discussion in this chapter has been focused on quantitative data that have numerical values. 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. (The next chapter describes this approach in some detail.) Alternatively, if they are best described by qualitative adjectives, such as dry or wet, hot or cold, expensive or cheap, clean or dirty, and high or low, membership functions indicating the fraction of stakeholders who believe particular quantitative descriptions of parameter or decision variable values are indeed hot, or cold, or clean or dirty, etc., can be used to quantify these qualitative descriptions. Both probability distributions and membership functions of these uncertain or qualitative variables can be included in quantitative models. This section introduces how qualitative variables can be included within models used for the preliminary screening of alternative water resources plans and management policies.

5.5.1   Linguistic Functions

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 their opposites. 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 quality and temperature of water is *good* for swimming. These qualitative, or so-called "fuzzy," statements convey information despite the imprecision of the italicized adjectives. The next section illustrates how these linguistic qualitative descriptors can be incorporated into optimization models using membership functions.

5.5.2 Membership Functions

Assume a set *A* of real or integer numbers ranging from 18 to 25. Thus *A* = [18, 25]. 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 called 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. Membership functions range 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 persons 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 function defining the interval or range of water temperatures suitable for swimming is shown in Fig. 5.10. 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.
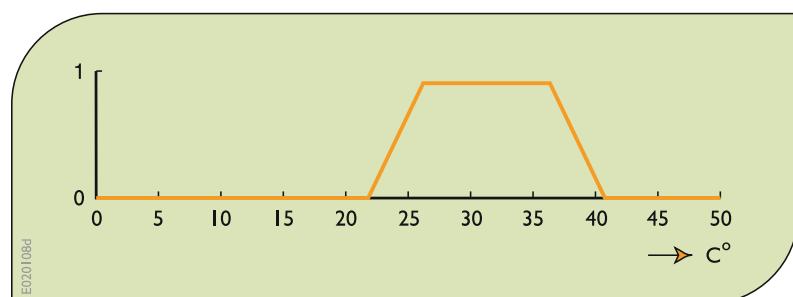
The form or shape of a function depends on the individual subjective feelings of the "members" or individuals who are asked their opinions. To define this particular function, each individual *i* could be asked to define his or her comfortable water temperature interval (T_{1i}, T_{2i}). The value associated with any temperature value *T* equals the number of individuals who place that *T* within their range (T_{1i}, T_{2i}), divided by the total number of individual opinions obtained. It is the fraction of the total number of individuals that consider the water temperature *T* suitable for swimming. For this reason such functions are often called membership functions (Figs. 5.10, 5.11 and 5.12).
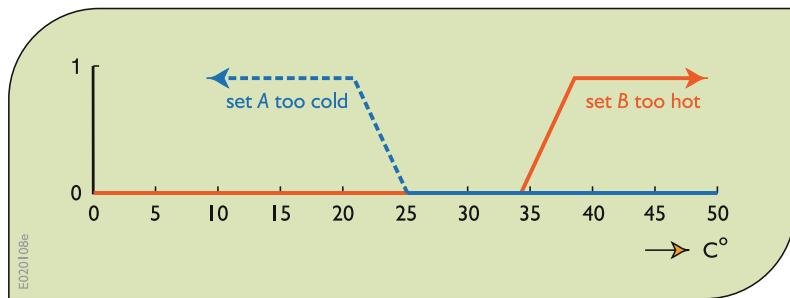
The assignment of membership values is based on subjective judgments, but such judgments seem to be sufficient for much of human communication.

Now suppose the water temperature applied to a swimming pool where the temperature could be regulated. The hotter the temperature the more it will cost. If we could quantify the economic benefits associated with various temperatures we could perform a benefit–cost analysis by maximizing the net benefits. Alternatively, we could maximize the fraction of people who consider the temperature good for swimming subject to a cost constraint using a membership function such as in Fig. 5.10 in place of an economic benefit function (Chap. 4 discusses ways of doing this.).

Continuing with this example, assume you are asked to provide the desired temperature at a reasonable cost. Just what is reasonable can also be defined by another membership function, but this time the function applies to cost, not temperature. Both the objective and constraint of this

Fig. 5.10 A membership function for suitability of water temperature for swimming
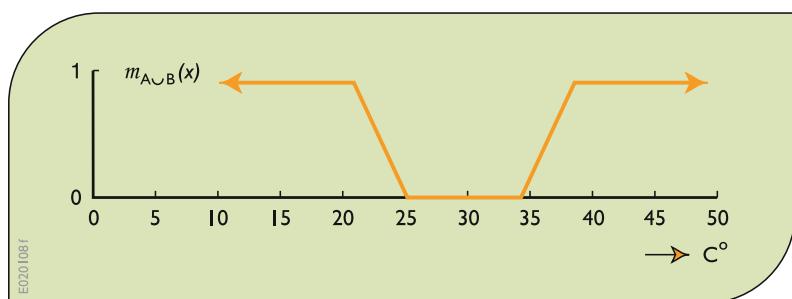
Fig. 5.11 Two membership functions relating to swimming water temperature. Set *A* is the set defining the fraction of all individuals who think the water temperature is too cold, and Set *B* defines the fraction of all individuals who think the water temperature is too hot

Fig. 5.12 Membership function for water temperatures that are considered too cold or too hot

problem are described qualitatively. In this case one could consider there are in fact two objectives, suitable temperature and acceptable cost. A model that maximizes the minimum value of both membership functions is one approach for finding an acceptable policy for controlling the water temperature at this swimming pool.

5.5.3 Illustrations of Qualitative Modeling

5.5.3.1 Water Allocation
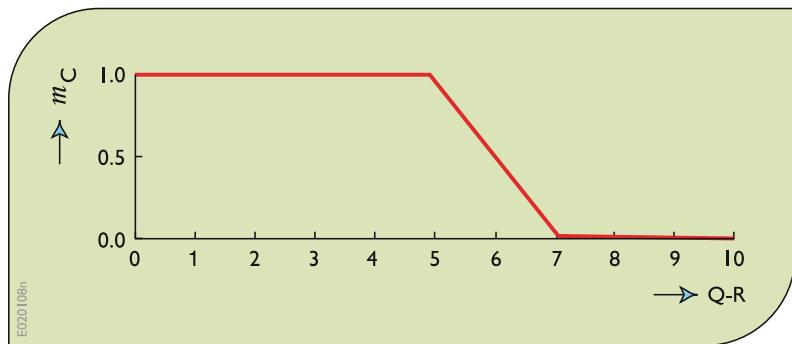
Consider the application of qualitative modeling to the water allocation problem illustrated in Fig. 5.7. Assume, as in the previous uses of this example, the problem is to find the allocations of water to each firm that maximize the total benefits TB(*X*):

$$\text{Maximize TB}(X) = \left(6x_1 - x_1^2\right) + \left(7x_2 - 1.5x_2^2\right) \\ + \left(8x_3 - 0.5x_3^2\right)$$

$$(5.10)$$

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, $Q - R$, as 6, the crisp optimization problem is to maximize Eq. (5.10) subject to the resource constraint:

$$x_1 + x_2 + x_3 \leq 6 \qquad (5.11)$$

The optimal solution is $x_1 = 1$, $x_2 = 1$, and $x_3 = 4$ as previously obtained in Chap. 4 using any of several different optimization methods. The maximum total benefit, TB(*X*), from Eq. (5.10), equals 34.5.

To create a qualitative 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.

Fig. 5.13 Membership
function for "about 6 units
more or less"

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

$$m(X) = \left[\left(6x_1 - x_1^2\right) + \left(7x_2 - 1.5x_2^2\right) \right. \\ \left. + \left(8x_3 - 0.5x_3^2\right)\right]/49.17 \qquad (5.12)$$

It is obvious that the two functions (Eqs. 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

$$\text{maximize } m(X) = \left[\left(6x_1 - x_1^2\right) + \left(7x_2 - 1.5x_2^2\right) \right. \\ \left. + \left(8x_3 - 0.5x_3^2\right)\right]/49.17$$
$$(5.13)$$

subject to

$$x_1 + x_2 + x_3 \leq 6 \qquad (5.14)$$

The optimal solution of (5.13) and (5.14)
results in the same values of each allocation as do
Eqs. (5.10) and (5.11). The optimal degree of
satisfaction is $m(X) = 0.70$.

Next, assume the total amount of resources
available to be allocated is limited to "about 6
units more or less," which is a qualitative

constraint. Assume the membership function
describing this constraint is defined by Eq. (5.14)
and is shown in Fig. 5.13.

$$m_C(X) = 1 \quad \text{if} \quad x_1 + x_2 + x_3 \leq 5$$
$$m_C(X) = [7 - (x_1 + x_2 + x_3)]/2 \quad \text{if} \quad 5 \leq x_1 + x_2 + x_3 \leq 7$$
$$m_C(X) = 0 \quad \text{if} \quad x_1 + x_2 + x_3 \geq 7$$
$$(5.15)$$

Let the membership function of (5.12) be
called $m_G(X)$. The qualitative optimization
problem becomes one of maximizing the mini-
mum value of the two membership functions
$(m_G(X), m_C(X))$ subject to their definitions in
Eqs. (5.12) and (5.15).

This yields $x_1 = 0.91$, $x_2 = 0.94$, $x_3 = 3.81$,
$m_G(X) = m_C(X) = 0.67$, and the total net benefit,
Eq. (5.10), is TB(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.

5.5.3.2 Qualitative Reservoir Storage and Release Targets

Consider the problem of trying to identify a
reservoir storage volume target, T^S, for 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 that are
less than the minimum release target.

Table 5.4 The solution to the reservoir optimization problem

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	16.9	reservoir release during period 1
R_2	37.5	reservoir release during period 2
R_3	20.6	reservoir release during period 3

Given a sequence of inflows, Q_t, the optimization model is

$$\text{Minimize } D = \sum_t \left[(T^s - S_t)^2 + \text{DR}^2 \right] - 0.001 T^S \tag{5.16}$$

subject to

$$S_t + Q_t - R_t = S_{t+1} \quad t = 1,2,3; \\ \text{if } t = 3, t+1 = 1 \tag{5.17}$$

$$S_t \leq K \quad t = 1,2,3 \tag{5.18}$$

$$R_t \geq T^R - \text{DR}_t \quad t = 1,2,3 \tag{5.19}$$

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, is listed in Table 5.4.

Now consider changing the objective function into maximizing the weighted degrees of "satisfying" the reservoir storage volume and release targets.

$$\text{Maximize} \quad \sum_t (w_S m_{St} + w_R m_{Rt}) \tag{5.20}$$
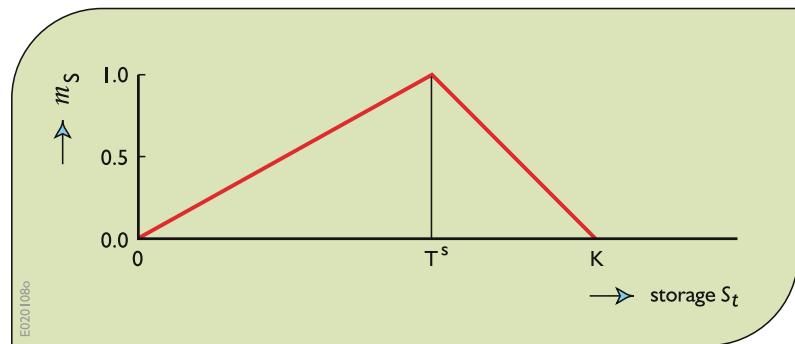
where w_S and w_R are weights indicating the relative importance of storage volume targets and release targets, respectively. The variables m_{St} are the degrees of satisfying the storage volume target in the three periods t, expressed by Eq. (5.21). The variables m_{Rt} are the degrees of satisfying the release target in periods t, expressed by Eq. (5.22).

$$m_{St} = S_t / \text{TS} \quad \text{for} \quad S_t \leq \text{TS} \quad \text{and} \\ (K - S_t)/(K - \text{TS}) \quad \text{for} \quad \text{TS} \leq S_t \tag{5.21}$$
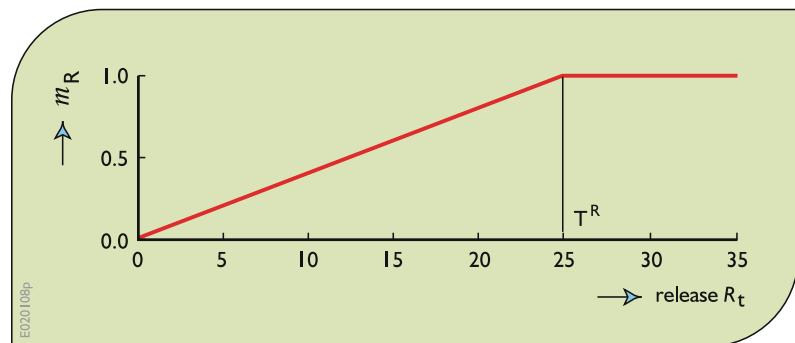
$$m_{Rt} = R_t / \text{TR} \ \text{ for } \ R_t \leq \text{TR} \quad \text{and} \\ 1 \ \text{ for } \ R_t > \text{TR} \tag{5.22}$$

Equations (5.21) and (5.22) are shown in Figs. 5.14 and 5.15, respectively.

Fig. 5.14 Membership
function for storage
volumes

Fig. 5.15 Membership
function for releases

Box 5.1. Reservoir model written for solution using LINGO.

```
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
```

Table 5.5 Solution of qualitative model for reservoir storage volumes and releases based on objective (5.20)

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

This optimization model 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 5.5.

If the objective Eq. 5.20 is changed to one of maximizing the minimum membership function value, the objective becomes

Maximize
$$m_{\min} = \text{maximize minimum}(m_{St}, m_{Rt})$$

(5.23)

To include the objective Eq. 5.23 in the optimization model a common lower bound, m_{min}, 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 5.6.

Table 5.6 Optimal solution of reservoir operation model based on objective (5.23)

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

This solution differs from that shown in Table 5.5 primarily in the values of the membership functions. The target storage volume operating variable value, T^s, stays the same value in this example.

5.5.3.3 Qualitative Water Quality Management Objectives and Constraints

Consider the stream pollution problem illustrated in Fig. 5.12. The stream receives waste, W_i from sources located at sites $i = 1$ 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 fraction of wastewater removal, x_i, 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 Fig. 5.16 are defined and listed in Table 5.7.

Using the notation defined in Table 5.7, the crisp model for this problem, as discussed in the previous chapter, is

$$\text{Minimize} \quad C_1(X_1) + C_2(X_2) \qquad (5.24)$$

subject to
Water quality constraint at site 2:

$$[P_1 Q_1 + W_1(1 - X_1)]a_{12}/Q_2 \le P_2^{\max}$$

$$[(32)(10) + 250{,}000(1 - X_1)/86.4]\,0.25/12$$
$$\le 20 \text{ which, when simplified, is}: X_1 \ge 0.78$$
$$(5.25)$$

Table 5.7 Parameter values selected for the water quality management problem illustrated in Fig. 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 3

Water quality constraint at site 3:

$$\{[P_1Q_1 + W_1(1 - X_1)]a_{13}$$
$$+ [W_2(1 - X_2)]a_{23}\}/Q_3 \leq P_3^{\max}$$
$$\{[(32)(10) + 250{,}000(1 - X_1)/86.4]\,0.15$$
$$+ [80{,}000(1 - X_2)/86.4]0.60\}/13 \leq 20 \tag{5.26}$$

which, when simplified, is: $X_1 + 1.28X_2 \geq 1.79$
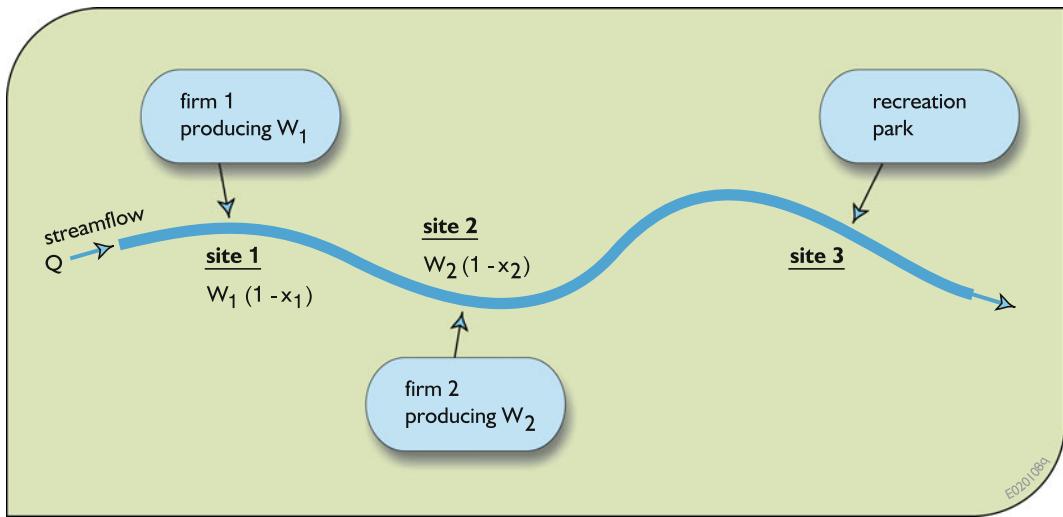
Restrictions on fractions of waste removal:

$$0 \leq X_i \leq 1.0 \quad \text{for sites } i = 1 \text{ and } 2 \tag{5.27}$$

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 qualitative model.
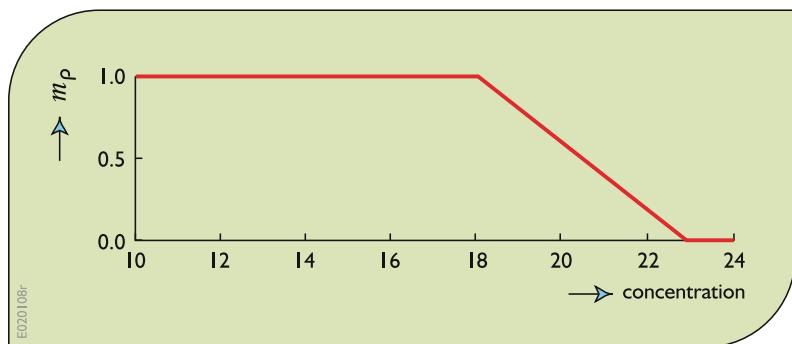
To consider a more qualitative version of this problem, suppose the maximum allowable pollutant concentrations in the stream at sites 2 and 3 were expressed as "about 20 mg/l more or less." Obtaining opinions of individuals of what they consider to be "20 mg/l more or less," a membership function can be defined. Assume it is as shown in Fig. 5.17.

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 Fig. 5.18.

Fig. 5.16 A stream pollution problem of finding the waste removal efficiencies (x_1, x_2) that meet the stream quality standards at least cost

Fig. 5.17 Membership function for "about 20 mg/l more or less"

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 Fig. 5.19.
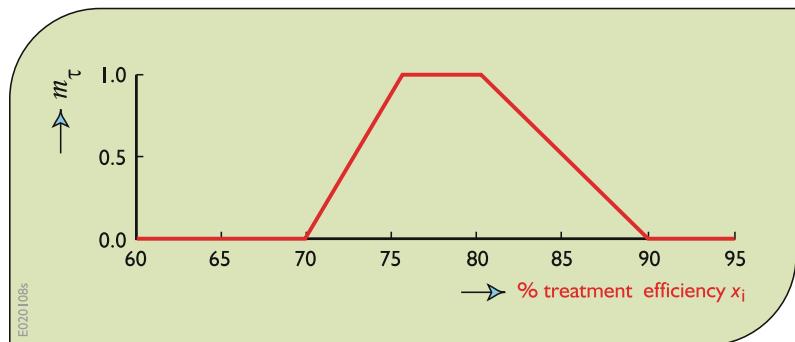
Considering each of these membership functions as objectives, a number of fuzzy optimization models can be defined. One is to find the treatment efficiencies that maximize the minimum value of each of these membership functions.
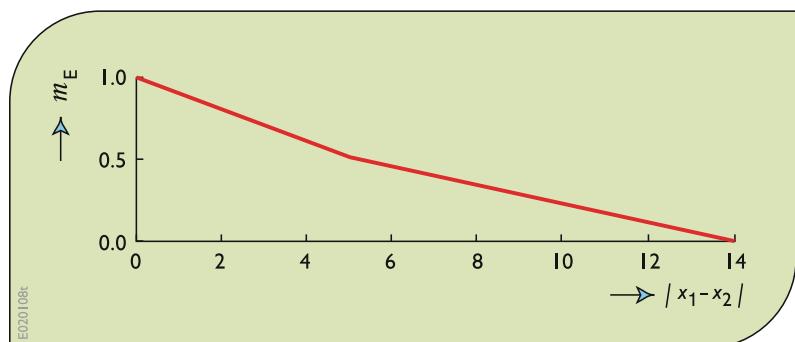
$$\text{Maximize} \quad m = \max \min\{m_P, m_T, m_E\} \tag{5.28}$$

If we assume that the pollutant concentrations at sites $j = 2$ and 3 will not exceed 23 mg/l, the

Fig. 5.18 Membership function for best available treatment technology

Fig. 5.19 Equity membership function in terms of the absolute difference between the two treatment efficiencies expressed as a percent

pollutant concentration membership functions m_{Pj} are

$$m_{Pj} = 1 - P_{2j}/5 \qquad (5.29)$$

The pollutant concentration at each site j is the sum of two components:

$$P_j = P_{1j} + P_{2j} \qquad (5.30)$$

where

$$P_{1j} \le 18 \qquad (5.31)$$

$$P_{2j} \le (23 - 18) \qquad (5.32)$$

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

$$m_{Ti} = (x_{2i}/0.05) - (x_{4i}/0.10) \qquad (5.33)$$

and the treatment efficiencies, expressed as fractions, are

$$X_i = 0.70 + x_{2i} + x_{3i} + x_{4i} \qquad (5.34)$$

where

$$x_{2i} \le 0.05 \qquad (5.35)$$

$$x_{3i} \le 0.05 \qquad (5.36)$$

$$x_{4i} \le 0.10 \qquad (5.37)$$

Finally, assuming the difference between treatment efficiencies will be no greater than 14, the equity membership function, m_E, is

$$m_E = Z - (0.5/0.05)D_1 + 0.5(1 - Z) \\ - (0.5/(0.14 - 0.05))D_2 \qquad (5.38)$$

where

$$D_1 \leq 0.05Z \qquad (5.39)$$

$$D_2 \leq (0.14 - 0.05)(1 - Z) \qquad (5.40)$$

$$X_1 - X_2 = DP - DM \qquad (5.41)$$

$$DP + DM = D_1 + 0.05(1 - Z) + D_2 \qquad (5.42)$$

$$Z \text{ is a binary } 0, 1 \text{ variable.} \qquad (5.43)$$

The remainder of the water quality model remains the same: Water quality constraint at site 2:

$$[P_1 Q_1 + W_1(1 - X_1)]a_{12}/Q_2 = P_2$$
$$[(32)(10) + 250{,}000(1 - X_1)/86.4]0.25/12 = P_2 \qquad (5.44)$$

Water quality constraint at site 3:

$$\{[P_1 Q_1 + W_1(1 - X_1)]a_{13} + [W_2(1 - X_2)]a_{23}\}/Q_3 = P_3$$
$$\{[(32)(10) + 250{,}000(1 - X_1)/86.4]\,0.15$$
$$+ [80{,}000(1 - X_2)/86.4]0.60\}/13 = P_3 \qquad (5.45)$$

Restrictions on fractions of waste removal:

$$0 \leq X_i \leq 1.0 \quad \text{for sites } i = 1 \text{ and } 2. \qquad (5.46)$$

Solving this model using LINGO® yields the results shown in Table 5.8.

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 linear programming (LP) model containing no membership functions.

Table 5.8 Solution to water quality management model Eqs. 5.28 to 5.46

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

5.6   Conclusions

Most computer-based models used for water resources planning and management are physical, mechanistic, or process-based quantitative 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 the system being modeled. 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 outlined some data-fitting methods of modeling that do not attempt to model natural, economic, or social processes. These have included ANN and two evolutionary search approaches: genetic algorithms (GA) for estimating the parameter and decision variable 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 system model, provide the most practical way to calibrate model parameters.

Fortunately for potential users of GA, GP, and ANN methods, software programs implementing many of these methods are available on the Internet. Applications of such methods to groundwater modeling, sedimentation processes along coasts and in harbors, 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.

Not all data are quantitative. In many cases objectives and constraints are expressed as qualitative expressions. Optimization models incorporating such expressions or functions are sometimes appropriate when only qualitative statements apply to a particular water management problem or issue. This chapter concludes by showing how this can be done using some simple example problems associated with water allocations, reservoir operation, and pollution control.

The information presented in this chapter serves only as an introduction. Those interested in more detailed and complete explanations and applications may refer to any of the additional references listed in the next section.

References

Aguilera, D. R. (2000). *Genetic programming with GPKER-NEL*. DHI, Copenhagen: Unpublished.

Babovic, V. (1996). Emergence, evolution, intelligence: Hydroinformatics. Delft: Delft University of Technology.

Babovic, V., & Abbott, M. B. (1997). The evolution of equations from hydraulic data. Part II: Applications. *Journal of Hydraulic Research, 35*(3), 411–427.

Babovic, V., & Keijzer, M. (2000). Genetic programming as a model induction engine. *Journal of Hydroinformatics, 2*(1), 35–60.

Banzhaf, W., Nordin, P., Keller, R. E., & Frannkone, F. D. (1998). *Genetic programming: An introduction: On the automatic evolution of computer programs and its applications.* San Francisco, California: Morgan Kaufmann Publishers, Inc. and dpunkt—Verlag fur Digitale Technologie GmbH.

Das, S., & Suganthan, P. N. (2011). Differential evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation*, *15*(1), 4–31.

Drecourt, J.-P. (1999). *Application of neural networks and genetic programming to rainfall-runoff modeling, D2K Technical Report 0699-1*. Copenhagen: DHI.

Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). *Artificial intelligence through simulated evolution.* Hoboken: Wiley.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison-Westley: Reading.

Hagan, M. T., Demuth, H. B., & Beale, M. (1996). *Neural network design.* Boston, Massachusetts: PWS.

Hansen, N. (2006). The CMA evolution strategy: A comparing review. In *Towards a new evolutionary computation. Advances on estimation of distribution algorithms* (pp. 1769–1776). Berlin: Springer.

Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd ed.). N.J., Prentice-Hall: Upper Saddle River.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation.* Reading, Massachusetts: Addison Wesley.

Holland, J. H. (1975). Adaptation in natural and artificial systems. Ann Arbor, Michigan: University of Michigan Press.

Igel, C., Hansen, N., & Roth, S. (2007). Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation, 15*(1), 1–28.

Koza, J. R. (1992). Genetic programming: On the programming of computers by means of natural selection (complex adaptive systems). Cambridge, Massachusetts: MIT Press.

Koza, J. R., Bennett, F. H., III, Andre, D., & Keane, M. A. (1999). *Genetic programming III: Darwinian invention and problem solving.* San Francisco, California: Morgan Kaufmann.

Liong, S.-Y., Khu, S. T., Babovic, V., Havnoe, K., Chan, W. T., & Phoon, K. K. (2000). *Construction of non-linear models in water resources with evolutionary computation.* Singapore: National University of Singapore.

Mitchell, T. M. (1997). *Machine learning.* New York: McGraw-Hill.

Nicklow, J. W., Reed, P. M., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., et al. (2010). State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management, 136*(4), 412–432.

Price, K. V., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution—A practical approach to global optimization.* Berlin Heidelberg: Springer-Verlag. ISBN 3540209506.

Principe, J. C., Euliano, N. R., & Lefebvre, W. C. (2000). *Neural and adaptive systems: Fundamentals through simulation.* New York: Wiley.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Stuttgart: Frommann-Holzboog.

Schwefel, H. P. (1981). *Numerical optimization of computer models.* Hoboken: Wiley. ISBN 10: 471099880 | ISBN-13: 9780471099888.

Schwefel, H. P. (1995). *Evolution and optimum seeking.* Hoboken: Wiley.

See, L., & Openshaw, S. (1999). Applying soft computing approaches to river level forecasting. *Hydrological Sciences Journal, 44*(5), 763–778.

Simon, D. (2013). *Evolutionary optimization algorithms.* Hoboken: Wiley.

Smith, M. (1993). *Neural networks for statistical modeling.* New York: Van Nostrand Reinhold.

Storn, R., & Price, K. (1997). Differential evolution: a simple and efficient heuristic for global optimization over continuous spaces. *J Global Optimiz, 11,* 341–359.

Additional References (Further Reading)

Abrahart, R. J., Kneale, P. E., & See, L. M. (Eds.). (2004). *Neural networks for hydrological modeling.* Leiden, The Netherlands: AA Balkema.

Bardossy, A., & Duckstein, L. (1995). Fuzzy rule- based modeling with applications to geoPhysical, biological, and engineering systems. Boca Raton, Florida: CRC Press.

Babovic, V. (1998). A data mining approach to time series modeling and forecasting. In V. Babovic & L. C. Larsen (Eds.), *Proceedings of the Third International Conference on Hydroinformatics (Hydroinformatics '98)* (pp. 847–856). Rotterdam: Balkema.

Babovic, V., & Bojkov, V. H. (2001). *D2K technical report D2K TR 0401-1.* Copenhagen: DHI.

Beale, R., & Jackson, T. (1990). *Neural computing: An introduction.* Bristol: IOP.

Chen, S. Y. (1994). *Theory and application of fuzzy system decision-making.* Dalian, China: Dalian University of Technology Press. (In Chinese.).

Clair, T. A., & 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, 34*(3), 447–455.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems, 2,* 303–314.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2,* 359–366.

Hsieh, W. W., & Tang, B. (1998). Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society, 79,* 1855–1870.

Kindler, J. (1992). Rationalizing water requirements with aid of fuzzy allocation model. *Journal of Water Resources Planning and Management, ASCE, 118*(3), 308–323.

Kundzewicz, W. (Ed.). (1995). *New uncertainty concepts in hydrology and water resources.* Cambridge, UK: Cambridge University Press.

Lange, N. T. G. (1998). Advantages of unit hydrograph derivation by neural networks. In V. Babovic & L. C. Larsen (Eds.), *Hydroinformatics '98* (pp. 783–789). Rotterdam: Balkema.

Lootsma, F. A. (1997). *Fuzzy logic for planning and decision-making.* Boston, Massachusetts: Kluwer Academic.

McKinney, D. C., & Lin, M. D. (1994). Genetic algorithm solution of groundwater management models. *Water Resources Research, 30*(6), 1897–1906.

Minns, A. W. (1996). Extended rainfall-runoff modeling using artificial neural networks. In A. Müller (Ed.), *Hydroinformatics '96* (pp. 207–213). Rotterdam: Balkema.

Minns, A. W., & Hall, M. J. (1996). Artificial neural networks as rainfall-runoff models. *Hydrological Sciences Journal, 41*(3), 399–417.

Olivera, R., & Loucks, D. P. (1997). Operating rules for multi-reservoir operation. *Water Resources Research, 33*(4), 839–852.

Price, R. K., Samedov, J. N., & Solomatine, D. P. (1998). An artificial neural network model of a generalized

channel network. In V. Babovic & L. C. Larsen (Eds.), *Hydroinformatics '98* (pp. 813–818). Rotterdam: Balkema.

Proano, C. O., Verwey, A., van den Boogaard, H. F. P., & Minns, A. W. (1998). Emulation of a sewerage system computational model for the statistical processing of large number of simulations. In V. Babovic & L. C. Larsen (Eds.), *Hydroinformatics '98* (Vol. 2, pp. 1145–1152). Rotterdam: Balkema.

Recknagel, F., French, M., Harkonen P., & Yanunaka, K. I. (1997). Artificial neural network approach for modeling and prediction of algal blooms. *Ecological Modeling, 96*, 11–28.

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., & Kollat, J. B. (2013). Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in Water Resources, 51*, 438–456, doi:10.1016/j.advwatres.2012.01.005

Sanchez, L., Arroyo, V., Garcia, J., Koev, K., & Revilla, J. (1998). Use of neural networks in design of coastal sewage systems. *Journal of Hydraulic Engineering, 124*(5), 457–464.

Scardi, M. (1996). Artificial neural networks as empirical models for estimating phytoplankton production. *Marine Ecology Progress Series, 139*, 289–299.

Shen, Y., Solomatine, D. P., & 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, 42*, 751–756.

Simonovic, S. P. (2009). *Managing water resources: Methods and tools for a systems approach* (pp. 576). Paris; London: UNESCO; Earthscan James & James. ISBN 978-1-84407-554-6.

Solomatine, D. P., & Avila Torres, L. A. (1996). Neural network application of a hydrodynamic model in optimizing reservoir operation. In: A. Müller (Ed.), *Hydroinformatics '96* (pp. 201–206). Rotterdam: Balkema.

Stelling, G. S. (2000). A numerical method for inundation simulations. In: Y. N. Yoon, B. H. Jun, B. H. Seoh, & G. W. Choi (Eds.), *Proceedings of the 4th International Conference on Hydro-Science and Engineering, Seoul*, September 26–26, 2000. Seoul: Korea Water Resources Association.

Terano, T., Asai, K., & Sugeno, M. (1992). *Fuzzy systems theory and its application*. San Diego, California: Academic Press.

Tilmant, A., Vancloster, M., Duckstein, L., & Persoons, E. (2002). Comparison of fuzzy and nonfuzzy optimal reservoir operating policies. *Journal of Water Resources Planning and Management, ASCE, 128*(6), 390–398.

van den Boogaard, H. F. P., Gautam, D. K., & Mynett, A. E. (1998). Auto-regressive neural networks for the modeling of time series. In V. Babovic & L. C. Larsen (Eds.), *Hydroinformatics '98* (pp. 741–748). Rotterdam: Balkema.

van den Boogaard, H. F. P., Ten Brummelhuis, P. G. J., & Mynett, A. E. (2000). On-line data assimilation in auto-regressive neural networks. In *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., & Van Den Boogaard, H. F. P. (1998). Neural network modeling of forces on vertical structures. In: *Proceedings of the 27th International Conference on Coastal Engineering,* pp. 2096–109. Reston, Virginia: ASCE press.

Wen, C.-G., & Lee, C.-S. (1998). A neural network approach to multiobjective optimization for water quality management in a river basin. *Water Resources Research, 34*(3), 427–436.

Wüst, J. C. (1995). Current prediction for shipping guidance. In B. Kappen & S. Gielen (Eds.), *Neural networks: Artificial intelligence and industrial applications* (pp. 366-73). *Proceedings of the 3rd Annual SNN Symposium on Neural Networks*, Nijmegen, The Netherlands, September 14–15, 1995. London: Springer-Verlag.

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, Massachusetts: Kluwer Academic.

Exercises

5.1 An upstream reservoir serves as a recreation site for swimmers, wind surfers, and boaters. It also serves as a flood storage reservoir in the second of four seasons or time periods in a year. The reservoir's releases can be diverted to an irrigation area. A wetland area further downstream receives the unallocated portion of the reservoir release plus the return flow from the irrigation area. The irrigation return flow contains a salinity concentration that can damage the ecosystem.

   (a) Assume there exist recreation lake level targets, irrigation allocation targets, and wetland flow and salinity targets. The challenge is to determine the reservoir releases and irrigation allocations so as to "best" meet these targets. This is the crisp' problem.

Data:

Reservoir storage capacity: 30 mcm;

During period 2 the flood storage capacity is 5 mcm;

Irrigation return flow fraction: 0.3 (i.e., 30% of that diverted for irrigation);

Salinity concentration of reservoir water: 1 ppt;

Salinity concentration of irrigation return flow: 20 ppt;

Reservoir average inflows for four seasons, respectively: 5, 50, 20, 10 mcm;

Targets for part (a):

Target maximum salinity concentration in wetland: 3 ppt;

Target storage target for all four seasons: 20 mcm;

Minimum flow target in wetland in each season, respectively: 10, 20, 15, 15 mcm;

Maximum flow target in wetland in each season, respectively: 20, 30, 25, 25 mcm;

Target irrigation allocations: 0, 20, 15, 5 mcm;

(b) Next create fuzzy membership functions to replace the targets and solve the problem. Assume that the targets used in (a) above are expressed in qualitative terms as membership functions. The membership functions indicate the degree of satisfaction for these targets. Solve for the "best" reservoir release and allocation policy that maximizes the minimum membership function value. Each membership function defines the relative level of satisfaction, where a value of 1 indicates complete stakeholder satisfaction. This is the qualitative problem.

5.2 Develop a flow chart showing how you would apply genetic algorithms to finding the parameters, a_{ij}, of a water quality prediction model, such as the one we have used to find the concentration downstream of an upstream discharge site i. This will be based on observed values of mass inputs, W_i, and concentrations, C_j, and flows, Q_j, at a downstream site j.

$$C_j = \sum_i W_i a_{ij}/Q_j$$

The objective to be used for fitness is to minimize the sum of the differences between the observed C_j and the computed C_j. To convert this to a maximization objective you could use something like the following:

$$\text{Max} 1/(1+D)$$

where

$$D \geq \left(C_j\text{obs}-C_j\text{calculated}\right)$$
$$D \geq \left(C_j\text{calculated}-C_j\text{obs.}\right)$$
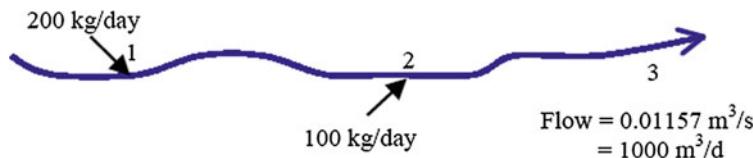
5.3 Use a genetic algorithm program to predict the parameter values asked for in problem 5.2, and then an artificial neural network ANN to obtain a predictor of downstream water quality based on the values of these parameters. You may use the model and data presented in Sect. 5.2 of Chap. 4 if you wish.

5.4. Using a genetic algorithm program to findthe allocations X_i that maximize the total benefits to the three water users i along a stream, whose individual benefits are

$$\text{Use 1:}\quad 6X_1 - X_1^2$$
$$\text{Use 2:}\quad 7X_2 - X_2^2$$
$$\text{Use 3:}\quad 8X_3 - X_3^2$$

Assume the available stream flow is some known value (ranging from 0 to 20).

Determine the effect of different genetic algorithm parameter values on the ability to find the best solution.

5.5 Consider the wastewater treatment problem illustrated in the drawing below.

200 kg/day

100 kg/day

Flow = 0.01157 m³/s
= 1000 m³/d

The initial stream concentration just upstream of site 1 is 32. The maximum concentration of the pollutant just upstream of site 2 is 20 mg/l (g/m^3), and at site 3 it is 25 mg/l. Assume the marginal cost per fraction (or percentage) of the waste load removed at site 1 is no less than that cost at site 2, regardless of the amount removed.

Using a suitable genetic algorithm program, solve for the least-cost wastewater treatment at sites 1 and 2 that will satisfy the quality constraints at sites 2 and 3, respectively.

Discuss the sensitivity of the GA parameter values in finding the best solution. You can get the exact solution using LINGO as discussed in Sect. 4.5.3.

5.6 Develop an artificial neural network for flow routing given the following two sets of upstream and downstream flows. Use one set of 5-periods for training (finding the unknown weights and other variables) and the other set for validation of the calculated parameter values (weights and bias constants).

Develop the simplest artificial neural network you can that does an adequate job of prediction.

Time period	Upstream flow	Downstream flow
1	450	366
2	685	593
3	830	755
4	580	636
5	200	325
1	550	439
2	255	304
3	830	678
4	680	679
5	470	534

[These outflows come from the following model, assuming an initial storage in period 1 of 50, the detention storage that will remain in the reach even if the inflows go to 0. For each period t:

$$\text{Outflow}(t) = 1.5(-50 + \text{initial storage}(t) + \text{inflow}(t))^{0.9},$$

where the outflow is the downstream flow and inflow is the upstream flow.]

An Introduction to Probability, Statistics, and Uncertainty

<div style="text-align: right">**6**</div>

Processes that are not fully understood, and whose outcomes cannot be precisely predicted, are often called uncertain. Most of the inputs to, and processes that occur in, and outputs resulting from water resource systems are not known with certainty. Hence so too are the future impacts of such systems, and even people's reactions and responses to these impacts. Ignoring this uncertainty when performing analyses in support of decisions involving the development and management of water resource systems could lead to incorrect conclusions, or at least more surprises, than will a more thorough analysis taking into account these uncertainties. This chapter introduces some commonly used approaches for dealing with model input and output uncertainty. Further chapters incorporate these tools in more detailed optimization, simulation, and statistical models designed to identify and evaluate alternative plans and policies for water resource system development and operation.

6.1 Introduction

Uncertainty is always present when planning and operating water resource systems. It arises because many factors that affect the performance of water resource systems are not and cannot be known with certainty when a system is planned, designed, built, and operated. The success and performance of each component of a system often depends on future meteorological, demographic, 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 (random over time) nature of meteorological and hydrological processes such as rainfall and evaporation. Similarly, future populations of towns and cities, per capita water usage rates, irrigation patterns, and priorities for water uses, all of which impact water demand, are not known with certainty. This chapter introduces methods for describing and dealing with uncertainty, and provides some simple examples of their use in water resources planning. These methods are extended in the following two chapters.

There are many ways to deal with uncertainty. The simplest approach is to replace each uncertain quantity either by its expected or average value or by some critical (e.g., "worst-case") value and then proceed with a deterministic approach. Use of *expected values* or alternatively *median values* of uncertain quantities can 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 values of uncertain parameters or variables are used in a deterministic model, the planner can then assess the importance of uncertainty with sensitivity and uncertainty analyses, discussed later in this and subsequent chapters.

Replacement of uncertain quantities by either expected or worst-case values can adversely affect the evaluation of project performance

Table 6.1 Data for determining reservoir recreation potential

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

when important parameters are highly variable. To illustrate these issues, consider the evaluation of the recreation potential of a reservoir. Table 6.1 shows that the elevation of the water surface varies from year to year 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 $\overline{L}$ is simply the sum of each possible pool level times its probability, or

$$\overline{L} = 10(0.10) + 20(0.25) + 30(0.30) \\ + 40(0.25) + 50(0.10) = 30 \qquad (6.1)$$

This pool level corresponds to 100 visitor-days per day

$$\mathrm{VD}(\overline{L}) = 100 \, \text{visitor-days per day} \qquad (6.2)$$

A worst-case analysis might select a pool level of 10 as a critical value, yielding an estimate of system performance equal to 25 visitor-days per day

$$\mathrm{VD}(L_{\mathrm{low}}) = \mathrm{VD}(10) = 25 \, \text{visitor-days per day} \qquad (6.3)$$

Neither of these values is a good approximation of the average visitation rate, which is

$$\overline{\mathrm{VD}} = 0.10\mathrm{VD}(10) + 0.25\mathrm{VD}(20) + 0.30\mathrm{VD}(30) \\ + 0.25\mathrm{VD}(40) + 0.10\mathrm{VD}(50) \\ = 0.10(25) + 0.25(75) + 0.30(100) \\ + 0.25(80) + 0.10(70) \\ = 78.25 \, \text{visitor-days per day}$$

$$(6.4)$$

Clearly, the average visitation rate, $\overline{\mathrm{V D}} = 78.25$, the visitation rate corresponding to the average pool level $\mathrm{VD}(\overline{L}) = 100$, and the worst-case assessment $\mathrm{VD}(L_{\mathrm{low}}) = 25$, are very different.

The median and the most likely are other measures that characterize a data set. They have the advantage that they are less influenced by extreme outliers. For the symmetric data set shown in Table 6.1, the median, most likely, and the mean are the same, namely 30. But if instead the probabilities of the respective pool levels were 0.30, 0.25, 0.20, 0.15, and 0.10, (instead of 0.10, 0.25, 0.30, 0.25, 0.10) the expected value or mean is 25, the value having the highest probability of occurring (the most likely) is 10, and the median or value that is greater or equal to half of the other values and less than or equal to the other half of the values in the data set is 20.

Thus 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, one should evaluate both the expected performance of a project and the risk and possible magnitude of project failures and their consequences.

This chapter reviews many of the methods of probability and statistics that are useful in water resources planning and management. Section 6.2 reviews the important concepts and methods of probability and statistics that are needed in this and other chapters of this book. Section 6.3 introduces several probability distributions that are often used to model or describe 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 6.4, 6.5, and 6.6 expand upon the use of these distributions, and discuss alternative parameter estimation methods.

Section 6.7 presents the basic ideas and concepts of 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 6.8 illustrates how synthetic flows and other time series inputs can be generated for stochastic simulations. The latter is introduced with an example in Sect. 6.9.

Many topics receive only brief treatment here. Readers desiring additional information should consult applied statistical texts such as Benjamin and Cornell (1970), Haan (1977), Kite (1988), Stedinger et al. (1993), Kottegoda and Rosso (1997), Ayyub and McCuen (2002), and Pishro-Nik (2014).

6.2 Probability Concepts and Methods

This section introduces basic concepts of probability and statistics. These are used throughout this chapter and in later chapters in the book.

6.2.1 Random Variables and Distributions

A basic concept in probability theory is that of the *random variable*. A random variable is a function whose value cannot be predicted with certainty. Examples of random variables 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 gage site on an unregulated stream. The values of all of these quantities depend on events that are not knowable before the event has occurred. Probability can be used to describe the likelihood 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 in a discrete set (such as the positive integers). The second two examples illustrate *continuous random variables*. Continuous random variables take on values in a continuous set. A property of all continuous random variables is that the probability that they equal any specific number is 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 can be non-zero. 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 probability that the value of X is less than or equal to a specific value or threshold x

$$F_X(x) = \Pr[X \le x] \tag{6.5}$$

This cumulative distribution function $F_X(x)$ is a non-decreasing function of x because

$$\Pr[X \le x] \le \Pr[X \le x + \delta] \quad \text{for } \delta > 0 \tag{6.6}$$

In addition,

$$\lim_{x \to +\infty} F_X(x) = 1 \tag{6.7}$$

and

$$\lim_{x \to -\infty} F_X(x) = 0 \tag{6.8}$$

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.

If X is a real-valued discrete random variable that takes on specific values x_1, x_2, ..., the *probability mass function* $p_X(x_i)$ is the probability X takes on the value x_i. Thus one would write

$$p_X(x_i) = \Pr[X = x_i] \tag{6.9}$$

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.

$$F_X(x) = \sum_{x_i \le x} p_X(x_i) \tag{6.10}$$

Figure 6.1 illustrates the probability mass function $p_X(x_i)$ and the cumulative distribution function of a discrete random variable.

The *probability density function* $f_X(x)$ for a continuous random variable X is the analogue of

the probability mass function of a discrete random variable. The probability density function, often called the pdf, is the derivative of the cumulative distribution function so that

$$f_X(x) = \frac{dF_X(x)}{dx} \ge 0 \tag{6.11}$$

The area under a probability density function always equals 1.

$$\int_{-\infty}^{+\infty} f_X(x) = 1 \tag{6.12}$$

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

$$\Pr[a < X \le b] = F_X(b) - F_X(a) = \int_a^b f_X(x)dx \tag{6.13}$$

The probability density function specifies the relative frequency with which the value of a continuous random variable falls in different intervals.

Life is seldom 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

$$F_{XY}(x, y) = \Pr[X \le x \text{ and } Y \le y]$$
$$= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v)dudv \tag{6.14}$$

If two random variables are discrete, then

$$F_{XY}(x, y) = \sum_{x_i \le x} \sum_{y_i \le y} p_{XY}(x_i, y_i) \tag{6.15}$$

where the joint probability mass function is

$$p_{XY}(x_i, y_i) = \Pr[X = x_i, \text{ and } Y = y_i] \quad (6.16)$$

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, the two random variables are said to be *independent*. Independence is an important and useful idea when attempting to develop a model of two or more random variables. For independent random variables

$$\Pr[a \le X \le b \text{ and } c \le Y \le d] \\ = \Pr[a \le X \le b]\Pr[c \le Y \le d] \quad (6.17)$$

for any a, b, c, and d. As a result,

$$F_{XY}(x, y) = F_X(x)F_Y(y) \quad (6.18)$$
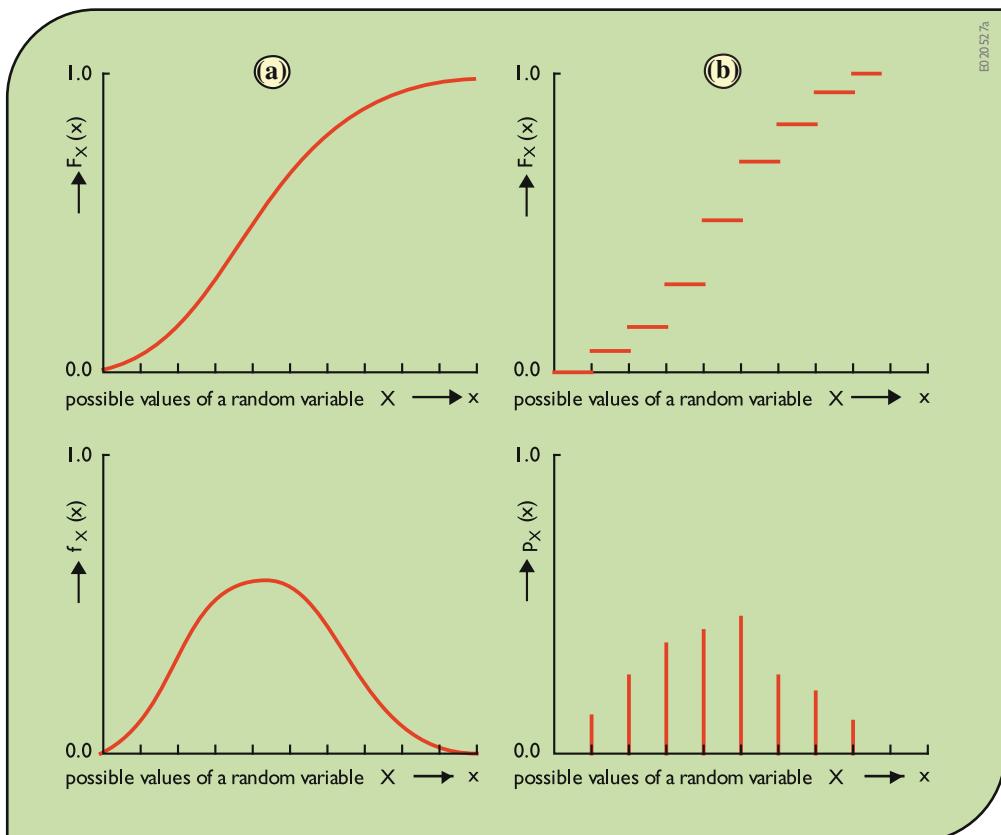
which implies for continuous random variables that

$$f_{XY}(x, y) = f_X(x)f_Y(y) \quad (6.19)$$

and for discrete random variables that

$$p_{XY}(x, y) = p_X(x)p_Y(y) \quad (6.20)$$

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

Fig. 6.1 Cumulative distribution and probability density or mass functions of random variables: **a** continuous distributions; **b** discrete distributions

$$F_X(x) = \Pr[X \le x] = \lim_{y \to \infty} F_{XY}(x, y) \quad (6.21)$$

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

$$f_X(x) = \int\limits_{-\infty}^{+\infty} f_{XY}(x, y)\mathrm{d}y \quad (6.22)$$

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

$$F_{X|Y}(x|y) = \Pr[X \le x | Y = y] = \frac{\int_{-\infty}^{x} f_{xy}(s, y)\mathrm{d}s}{f_y(y)} \quad (6.23)$$

It follows that the conditional density function is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} \quad (6.24)$$

For discrete random variables, the probability of observing $X = x$, given that $Y = y$ equals

$$p_{X|Y}(x|y) = \frac{p_{xy}(x, y)}{p_y(y)} \quad (6.25)$$

These results can be extended to more than two random variables. See Kottegoda and Rosso (1997) for a more advanced discussion.

6.2.2 Expected Values

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

$$E[g(X)] = \int\limits_{-\infty}^{+\infty} g(x)f_X(x)\mathrm{d}x \quad (6.26)$$

whereas for a discrete random variable

$$E[g(X)] = \sum_{x_i} g(x_i)p_X(x_i) \quad (6.27)$$

$E[\;]$ is called the *expectation operator*. It 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 nonrandom constants,

$$E[a + bX] = a + bE[X] \quad (6.28)$$

The expectation of a function of two random variables is given by

$$E[g(X, Y)] = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} g(x, y)f_{XY}(x, y)\mathrm{d}x\,\mathrm{d}y$$

or

$$E[g(X, Y)] = \sum_i \sum_j g(x_i, y_i)p_{XY}(x_i, y_i) \quad (6.29)$$

If X and Y are independent, the expectation of the product of a function $h(\cdot)$ of X and a function $g(\cdot)$ of Y is the product of the expectations

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)] \quad (6.30)$$

This follows from substitution of Eqs. 6.19 and 6.20 into Eq. 6.29.

6.2.3 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 pth 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

$$\Pr[X < x_p] \leq p \leq \Pr[X \leq x_p] \qquad (6.31)$$

Equation 6.31 is written to insist if at some point 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 pth 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

$$F_X(x_p) = p \qquad (6.32)$$

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 pth 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 often the selected design value. In water quality management, a river's minimum seven-day-average low flow expected once in 10 years is often used as the critical planning value: Here the one-in-ten year value is the 10 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 observed observations ordered by magnitude, and for a nonnegative integer k such that $n = 2k$ (even) or $n = 2k + 1$ (odd), the sample estimate of the median is

$$\hat{x}_{0.50} = \begin{cases} x_{(k+1)} & \text{for } n = 2k+1 \\ \frac{1}{2}\left[x_{(k)} + x_{(k+1)}\right] & \text{for } n = 2k \end{cases}$$
$$(6.33)$$

Sample estimates of other quantiles may be obtained 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 Sects. 6.3.1 and 6.3.2, and then estimate x_p using Eq. 6.32, where $F_X(x)$ is the fitted distribution.

Another simple and common approach to describing a distribution's center, spread, and shape is by reporting the moments of a distribution. The first moment about the origin μ_X is the *mean* of X and is given by

$$\mu_X = E[X] = \int_{-\infty}^{+\infty} x f_X(x)\,\mathrm{d}x \qquad (6.34)$$

Moments other than the first are normally measured about the mean. The second moment measured about the mean is the *variance*, denoted $\mathrm{Var}(X)$ or σ_X^2, where

$$\sigma_X^2 = \mathrm{Var}(X) = E[(X - \mu_X)^2] \qquad (6.35)$$

The *standard deviation* σ_X is the square root of the variance. While the mean μ_X is a measure of the central value of X, the standard deviation σ_X is a measure of the spread of the distribution of X about its mean μ_X.

Another measure of the variability in X is the *coefficient of variation*,

$$\text{CV}_X = \frac{\sigma_X}{\mu_X} \qquad (6.36)$$

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, which are both strictly positive values.

The third moment about the mean denoted λ_X, measures the asymmetry or *skewness* of the distribution

$$\lambda_X = E[(X - \mu_X)^3] \qquad (6.37)$$

Typically, the dimensionless coefficient of skewness γ_X is reported rather than the third moment λ_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

$$\gamma_X = \frac{\lambda_X}{\sigma_X^3} \qquad (6.38)$$

Streamflows and other natural phenomena that are necessarily nonnegative 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

$$\overline{X} = \sum_{i=1}^{n} X_i / n \qquad (6.39a)$$

The sample estimate of the variance

$$\hat{\sigma}_X^2 = S_X^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \overline{X})^2 \qquad (6.39b)$$

The sample estimate of skewness

$$\hat{\lambda}_X = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (X_i - \overline{X})^3 \qquad (6.39c)$$

The sample estimate of the coefficient of variation

$$\widehat{\text{CV}}_X = S_X / \overline{X} \qquad (6.39d)$$

The sample estimate of the coefficient of skewness

$$\hat{\gamma}_X = \hat{\lambda}_X / S_X^3 \qquad (6.39e)$$

The sample estimate of the mean and variance are often denoted $\bar{x}$ and s_X^2. All of these sample estimators only provide estimates. Unless the sample size n is very large, the difference between the estimators from the true values of $\mu_X, \sigma_X^2, \lambda_X$, CV_X, and γ_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 20 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 6.2. These data have the following sample moments:

$$\bar{x} = 1549.2$$
$$s_X = 813.5$$
$$\widehat{\text{CV}}_X = 0.525$$
$$\hat{\gamma}_X = 0.712$$

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 θ is a function of the values of the random variable $X_1, \ldots, X_n$ that will be available to estimate the value of θ; $\hat{\theta}$ may be written $\hat{\theta}[X_1, X_2, \ldots, X_n]$ to emphasize that $\hat{\theta}$

Table 6.2   Annual Maximum Discharges on Magra River, Italy, at Calamazza, 1930–1970

date	discharge cu ft/s	date	discharge cu ft/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

The value for 1945 ismissing

itself is a random variable because its value depends on the sample values of the random variable that will be observed. An estimator $\hat{\theta}$ of a quantity θ 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 μ_X because

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu_X$$

(6.40)

The estimator S_X^2 of the variance of X is an unbiased estimator of the true variance σ_X^2 for independent observations (Benjamin and Cornell 1970):

$$E[S_X^2] = \sigma_X^2 \qquad (6.41)$$

However, the corresponding estimator of the standard deviation, S_X, is in general a biased estimator of σ_x because

$$E[S_X] \neq \sigma_X \qquad (6.42)$$

The second important statistic often used to assess the accuracy of an estimator $\hat{\theta}$ is the

variance of the estimator $\text{Var}(\hat{\theta})$, 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

$$\text{Var}(\overline{X}) = \frac{\sigma_{\overline{X}}^2}{n} \qquad (6.43)$$

It is common to call $\sigma_x/\sqrt{n}$ the *standard error* of $\overline{X}$ rather than its standard deviation. The standard error of an average is the most commonly reported measures of the 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 θ 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 the two

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E\left[\left(\hat{\theta} - \theta\right)^2\right] \\ &= \left\{E[\hat{\theta}] - \theta\right\}^2 + E\left\{\left(\hat{\theta} - E[\hat{\theta}]\right)^2\right\} \\ &= [\text{Bias}]^2 + \text{Var}(\hat{\theta}) \end{aligned}$$
$$(6.44)$$

where [Bias] is $E(\hat{\theta}) - \theta$. Equation 6.44 shows that the MSE, equal to the expected average squared deviation of the estimator $\hat{\theta}$ from the true value of the parameter θ, can be computed as the bias squared plus the variance of the estimator. MSE is a convenient measure of how closely $\hat{\theta}$ approximates θ because it combines both bias and variance in a logical way.

Estimation of the coefficient of skewness γ_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 γ_X is often biased, has a large variance, and was shown by Kirby (1974) to be bounded so that

$$|\hat{\gamma}_X| \leq \sqrt{n} \qquad (6.45)$$

where n is the sample size. The bounds do not depend on the true skew γ_X. However, the bias and variance of $\hat{\gamma}_X$ do depend on the sample size and the actual distribution of X. Table 6.3 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. (1974a, b) who employed moment estimators slightly different than those in Eq. 6.39a.

For the normal distribution, $E[\hat{\gamma}_X] = 0$ 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 6.3 it is also 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 squared. Using Eq. 6.44, the mean square error of $\hat{\gamma}_X$ is

$$\begin{aligned} \text{MSE}(\hat{\gamma}_X) &= (0.45 - 0.50)^2 + (0.37)^2 \\ &= 0.0025 + 0.1369 = 0.139 \cong 0.14 \end{aligned}$$
$$(6.46)$$

An unbiased estimate of γ_X is simply $(0.50/0.45)\hat{\gamma}_X$. Here the estimator provided by Eq. 6.39a has been scaled to eliminate bias. This unbiased estimator has mean squared error

$$\begin{aligned} \text{MSE}\left(\frac{0.50\hat{\gamma}_x}{0.45}\right) &= (0.50 - 0.50)^2 + \left[\left(\frac{0.50}{0.45}\right)(0.37)\right]^2 \\ &= 0.169 \cong 0.17 \end{aligned}$$
$$(6.47)$$

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 6.3 except for the normal distribution for

Table 6.3 Sampling properties of coefficient of skewness estimator

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

Source Wallis et al. (1974b) who divided just by n in the estimators of the moments, whereas in Eqs. 6.39b and 6.39c we use the generally adopted coefficients of $1/(n-1)$ and $n/(n-1)(n-2)$ for the variance and skew

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).

6.2.4 L-Moments and Their Estimators

L-moments are another way to summarize the statistical properties of hydrologic data based on linear combinations of the original sample (Hosking 1990). Recently, hydrologists have found that regionalization methods (to be discussed in Sect. 6.5) 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).

The first L-moment designated as λ_1 is simply the arithmetic mean

$$\lambda_1 = E[X] \tag{6.48}$$

Now let $X_{(i|n)}$ be the ith largest observation in a sample of size n ($i = n$ corresponds to the largest). Then, for any distribution, the second L-moment, λ_2, is a description of scale based upon the expected difference between two randomly selected observations.

$$\lambda_2 = (1/2)E\big[X_{(2|1)} - X_{(1|2)}\big] \tag{6.49}$$

Similarly, L-moment measures of skewness and kurtosis use three and four randomly selected observations, respectively.

$$\lambda_3 = (1/3)E\big[X_{(3|3)} - 2X_{(2|3)} + X_{(1|3)}\big] \tag{6.50}$$

$$\lambda_4 = (1/4)E\big[X_{(4|4)} - 3X_{(3|4)} + 3X_{(2|4)} - X_{(1|4)}\big] \tag{6.51}$$

Sample estimates are often computed using intermediate statistics called probability weighted moments (PWMs). The rth probability weighted moment is defined as

$$\beta_r = E\{X[F(X)]^r\} \tag{6.52}$$

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 β_r are computed as

$$b_0 = \overline{X}$$
$$b_1 = \frac{1}{n(n-1)}\sum_{j=2}^{n}(j-1)X_{(j)}$$
$$b_2 = \frac{1}{n(n-1)(n-2)}\sum_{j=3}^{n}(j-1)(j-2)X_{(j)} \tag{6.53}$$

These are examples of the general formula for computing estimators b_r of β_r.

$$b_r = \frac{1}{n}\sum_{i=r}^{n}\binom{i}{r}X_{(i)}\bigg/\binom{n-1}{r}$$
$$= \frac{1}{r+1}\sum_{i=r}^{n}\binom{i}{r}X_{(i)}\bigg/\binom{n}{r+1} \tag{6.54}$$

for $r = 1, \ldots, n-1$.

L-moments are easily calculated in terms of probability weighted moments (PWMs) using

$$\begin{aligned}\lambda_1 &= \beta_0 \\ \lambda_2 &= 2\beta_1 - \beta_0 \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0\end{aligned} \tag{6.55}$$

Formulas for directly calculating L-moment estimators, b, of β, are provided by Wang (1997). 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. Whereas 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

Table 6.4 Definitions of dimensionless product-moment and L-moment ratios

name	common symbol	definition
product-moment ratios		
coefficient of variation	CV_X	σ_X/μ_X
skewness	γ_X	$E[(X-\mu_X)^3]/\sigma_X^3$
kurtosis	κ_X	$E[(X-\mu_X)^4]/\sigma_X^4$
L-moment ratios		
L-coefficient of variation *	L-CV, $\quad \tau_2$	λ_2/λ_1
skewness	L-skewness, τ_3	λ_3/λ_2
kurtosis	L-kurtosis, $\quad \tau_4$	λ_4/λ_2

* Hosking and Wallis (1997) use τ instead of τ_2 to represent the L-CV ratio

zero. Table 6.4 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 6.2 can be used to provide an example of L-moments. Equation 6.53 yields estimates of the first three Probability Weighted Moments

$$
\begin{aligned}
b_0 &= 1549.20 \\
b_1 &= 1003.89 \\
b_2 &= 759.02
\end{aligned}
\tag{6.56}
$$

Recall b_0 is just the sample average $\bar{x}$. The sample L-moments are easily calculated using the probability weighted moments (PWMs). One obtains

$$
\begin{aligned}
\hat{\lambda}_1 &= b_0 = 1549 \\
\hat{\lambda}_2 &= 2b_1 - b_0 = 458 \\
\hat{\lambda}_3 &= 6b_2 - 6b_1 + b_0 = 80
\end{aligned}
\tag{6.57}
$$

Thus the sample estimates of the L-Coefficient of Variation, t_2, and L-Coefficient of Skewness, t_3, are

$$
\begin{aligned}
t_2 &= 0.295 \\
t_3 &= 0.174
\end{aligned}
\tag{6.58}
$$

6.3   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 that one 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 6-h-rainfall depth, or the minimum seven-day-average expected once-in-10-year flow. Then the fitted distribution and its statistical parameters can characterize that random variable. 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 does have several advantages over use of the empirical distribution

1. 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 easily estimated compared to those computed from the empirical distribution.
2. It provides a compact and easy to use representation of the data.
3. 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, using the empirical distribution one often assumes that no values larger or smaller than the sample maximum or minimum can occur. For many situations this is unreasonable.
4. Often one needs to estimate the likelihood of extreme events that lie outside of 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.
5. In many cases one is not interested in X, but instead is interested in derived 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 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 determining if a fitted distribution provides a reasonable or acceptable model of the data. Subsections 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 resource planning (Kottegoda and Rosso 1997).

6.3.1 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 observed data could have been drawn from the fitted distribution. If not, the fitted distribution is judged to be unacceptable.

In many cases, good estimates of a distribution's parameters are obtained by the *maximum likelihood-estimation* procedure. Given a set of *n* independent observations $\{x_1, \ldots, x_n\}$ of a continuous random variable X, the joint probability density function for the observations is

$$
\begin{aligned}
f_{X_1, X_2, X_3, \ldots X_n}&(x_1, \ldots, x_n | \theta) \\
&= f_X(x_1|\theta) \cdot f_X(x_2|\theta) \ldots f_X(x_n|\theta)
\end{aligned} \tag{6.59}
$$

where θ is the vector of the distribution's parameters.

The maximum likelihood estimator of θ is that vector which maximizes Eq. 6.59 and thereby makes it as likely as possible to have observed the values $\{x_1, \ldots, x_n\}$.

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 to be highly 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

$$
f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} [\ln(x) - \mu]^2 \right\} \tag{6.60}
$$

Here the parameters μ and σ^2 are the mean and variance of the logarithm of X, and not of X itself.

Maximizing the logarithm of the joint density for $\{x_1, \ldots, x_n\}$ is more convenient than maximizing the joint probability density itself. Hence the problem can be expressed as the maximization of the *log-likelihood function*

$$
\begin{aligned}
L &= \ln\left[\prod_{i=1}^{n} f(x_i|\mu, \sigma) \right] \\
&= \sum_{i=1}^{n} \ln f(x_i|\mu, \sigma) \\
&= -\sum_{i=1}^{n} \ln\left(x_i\sqrt{2\pi} \right) \\
&\quad - n\ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} [\ln(x_i) - \mu]^2
\end{aligned} \tag{6.61}
$$

The maximum can be obtained by equating to zero the partial derivatives $\partial L/\partial\mu$ and $\partial L/\partial\sigma$ whereby one obtains

$$
\begin{aligned}
0 &= \frac{\partial L}{\partial\mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} [\ln(x_i) - \mu] \\
0 &= \frac{\partial L}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} [\ln(x_i) - \mu]^2
\end{aligned} \tag{6.62}
$$

These equations yield the estimators

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\ln(x_i)$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}[\ln(x_i) - \hat{\mu}]^2 \qquad (6.63)$$

The second-order conditions for a maximum are met and these values do maximize Eq. 6.59. It is useful to note that if one defines a new random variable $Y = \ln(X)$, then the maximum likelihood estimates of the parameters μ and σ^2, which are the mean and variance of the Y distribution, are the sample estimates of the mean and variance of Y

$$\hat{\mu} = \bar{y}$$

$$\hat{\sigma}^2 = [(n-1)/n]s_Y^2 \qquad (6.64)$$

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 postulated distribution in Eqs. 6.34, 6.35, and 6.37 are equated to the estimates of those moments calculated using Eqs. 6.39a. 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 μ and σ are given by

$$\mu_X = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

$$\sigma_X^2 = \exp(2\mu + \sigma^2)\left[\exp(\sigma^2) - 1\right] \qquad (6.65)$$

Substituting $\bar{x}$ for μ_X and s_x^2 for σ_X^2 and solving for μ and σ^2 one obtains

$$\hat{\sigma}^2 = \ln(1 + s_X^2/\bar{x}^2)$$

$$\hat{\mu} = \ln\left(\frac{\bar{x}}{\sqrt{1 + s_X^2/\bar{x}^2}}\right) = \ln\bar{x} - \frac{1}{2}\hat{\sigma}^2 \qquad (6.66)$$

The data in Table 6.2 provide an illustration of both fitting methods. One can easily compute the sample mean and variance of the logarithms of the flows to obtain

$$\hat{\mu} = 7.202$$

$$\hat{\sigma}^2 = 0.3164 = (0.5625)^2 \qquad (6.67)$$

Alternatively, the sample mean and variance of the flows themselves are

$$\bar{x} = 1549.2$$

$$s_X^2 = 661{,}800 = (813.5)^2 \qquad (6.68)$$

Substituting those two values in Eq. 6.66 yields

$$\hat{\mu} = 7.224$$

$$\hat{\sigma}^2 = 0.2435 = (0.4935)^2 \qquad (6.69)$$

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 hydrologic literature (Landwehr et al. 1978; Hosking et al. 1985; 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 Sect. 6.5 (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

$$f_\theta(\boldsymbol{\theta}|x_1, x_2, \ldots, x_n) \propto f_X(x_1, x_2, \ldots, x_n|\boldsymbol{\theta})\xi(\boldsymbol{\theta})$$

$$(6.70)$$

The symbol $\propto$ means "proportional to" and $\xi(\boldsymbol{\theta})$ is the probability density function for the prior distribution for $\boldsymbol{\theta}$ (Kottegoda and Rosso 1997). Thus, except for a constant of proportionality, the probability density function describing the posterior distribution of the parameter vector $\boldsymbol{\theta}$ is equal to the product of the likelihood function $f_X(x_1, x_2, \ldots, x_n|\boldsymbol{\theta})$ and the probability density function for the prior distribution $\xi(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$.

Advantages of the Bayesian approach are that it allows the explicit modeling of uncertainty in parameters (Stedinger 1997; Kuczera 1999), and provides a theoretically consistent framework for integrating systematic flow records with regional and other hydrologic information (Vicens et al. 1975; Stedinger 1983; and 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 κ of a GEV distribution allowed definition of generalized maximum likelihood estimators that over the κ-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 Eq. 6.70 has been an obstacle to their use. Solutions to Eq. 6.70 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 allow 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 modeling, 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 Sect. 6.4 (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 a fitted curve is or is not 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 (Stedinger 1997). 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).

6.3.2   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 ith 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 pth quantile x_p of the true distribution of X for $p = i/(n + 1)$. In fact, if one thinks of the cumulative probability U_i associated with the random variable $X_{(i)}$, $U_i = F_X(X_{(i)})$, then if the observations $X_{(i)}$ are independent, the U_i have a beta distribution (Gumbel 1958) with probability density function

$$f_{U_i}(u) = \frac{n!}{(i-1)!(n-1)!} u^{i-1}(1-u)^{n-i}$$
$$0 \leq u \leq 1$$
$$(6.71)$$

This beta distribution has mean and variance of

$$E[U_i] = \frac{i}{n+1} \qquad (6.72a)$$

and

$$\mathrm{Var}(U_i) = \frac{i(n-i+1)}{(n+1)^2(n+2)} \qquad (6.72b)$$

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 $G(x)$ exactly equaled the true X-distribution $F_X[x]$, the plotted points will not fall exactly on a 45-degree line through the origin of the graph. This would only occur if $F_X[x_{(i)}]$ exactly equaled $i/(n+1)$ and therefore each $x_{(i)}$ exactly equaled $F_X^{-1}[i/(n+1)]$.

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 quantiles of the distribution of U_i obtained from integrating the probability density function in Eq. 6.71. The required incomplete beta function is also available in many software packages, including Microsoft Excel. Stedinger et al. (1993) report 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 those values.

Figure 6.2a, b 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 6.2 for the Magra River, Italy, at Calamazza for the years 1930–1970. The observations of $X_{(i)}$, given in Table 6.2, 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 formula 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 in 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

$$\Pr\left[G^{-1}\left(\frac{i}{n} - C_\alpha\right) \le X_{(i)} \le G^{-1}\left(\frac{i-1}{n} + C_\alpha\right) \text{for every i}\right]$$
$$= 1 - \alpha$$

(6.73)

where C_α is the critical value of the test at significance level α. Formulas for C_α as a function of n are contained in Table 6.5 for three cases: (1) when G is completely 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 GEV distribution with known shape parameter κ. For other distributions, the values obtained from Table 6.5 may be used to construct approximate simultaneous confidence intervals for every $X_{(i)}$.
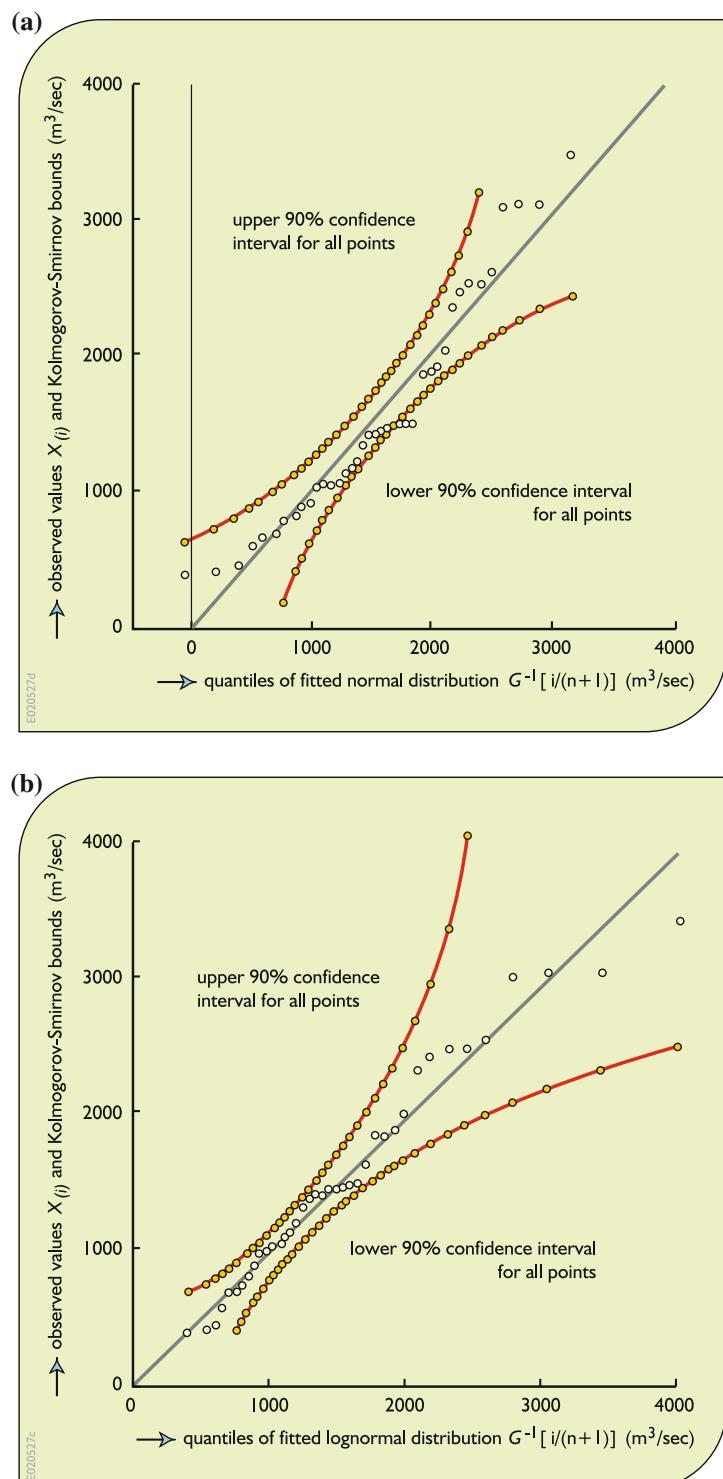
Figures 6.2 contain 90% confidence intervals for the plotted points constructed in this manner. For the normal distribution, the critical value of C_α equals $0.819/(\sqrt{n} − 0.01 + 0.85/\sqrt{n})$,

Table 6.5 Critical values[a] of Kolmogorov-Smirnov statistic as a function of sample size n

	significacne level α				
	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_α are calculated as follows:
for case 2 with $\alpha = 0.10$, $C_\alpha = 0.819 / (\sqrt{n} - 0.01 + 0.85 / \sqrt{n})$

Fig. 6.2 Plots of annual maximum discharges of Magra River, Italy, versus quantiles of fitted **a** normal and **b** lognormal distributions

where 0.819 corresponds to $\alpha = 0.10$. For $n = 40$, one computes $C_\alpha = 0.127$. As can be seen in Fig. 6.2a, 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 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 Fig. 6.2b, 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-degree 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 evaluating from which of several families a set of observations is likely to have been drawn. 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

$$r = \frac{\sum \left(x_{(i)} - \bar{x}\right)(w_i - \bar{w})}{\left[\left(\sum \left(x_{(i)} - \bar{x}\right)^2 \sum \left(w_i - \bar{w}\right)^2\right)\right]^{0.5}} \quad (6.74)$$

Table 6.6 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 6.7 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'Agostion and Stephens 1986).

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 $\text{Var}[\hat{\tau}_3] = (0.1866 + 0.8/n)/n$, allowing construction of a powerful test of normality against skewed alternatives using the normally distributed statistic

$$Z = t_3 / \sqrt{(0.1866 + 0.8/n)/n} \quad (6.75)$$

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 κ 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 κ, or a regional κ and 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

Table 6.6 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)

| 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

$$Z = (\hat{\tau}_3 - 0.17)/\sqrt{(0.2326 + 0.70/n)/n} \quad (6.76)$$

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'Agostion and Stephens 1986). Generation of a probability plot is most often a good idea because it allows the modeler to see what the data look like and where problems occur. The Kolmogorov–Smirnov test helps the eye 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.

6.3.3 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 density function of a normal random variable is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right] \quad (6.77)$$
$$\text{for} -\infty < x < +\infty$$

Table 6.7 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)

	significance level		
n	**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

where μ and σ^2 are equivalent to μ_X and σ_X^2, the mean and variance of X. Interestingly, the maximum likelihood estimators of μ and σ^2 are almost identical to the moment estimates $\bar{x}$ and s_X^2.

The normal distribution is symmetric about its mean μ_X and admits values from $-\infty$ to $+\infty$. Thus it is not always satisfactory for modeling physical phenomena such as streamflows or pollutant concentrations, which are necessarily nonnegative 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

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}[\ell n(x) - \mu]^2 \right\} \frac{d(\ell nx)}{dx}$$
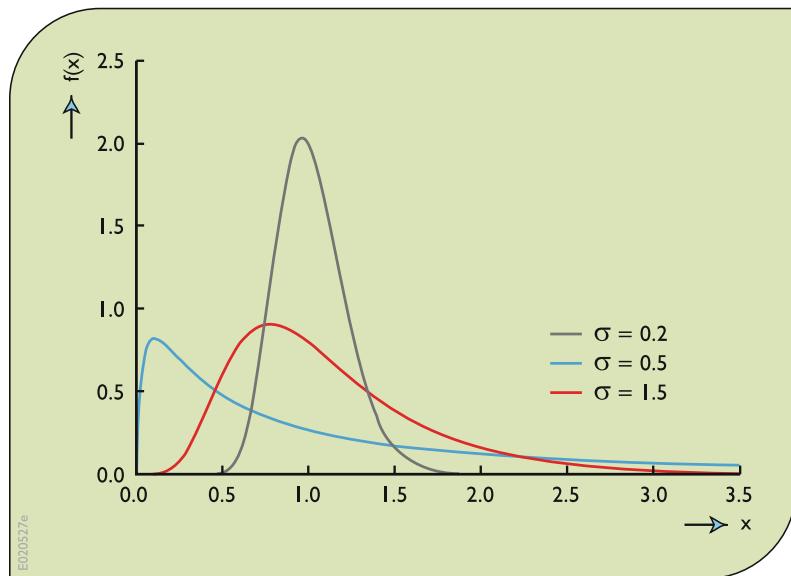
$$= \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}[\ell n(x/\eta)]^2 \right\}$$

(6.78)

for $x > 0$ and $\mu = \ln(\eta)$. Here η is the median of the X-distribution. The coefficient of skewness for the three-parameter lognormal distribution is given by

$$\gamma = 3\upsilon + \upsilon^3 \quad \text{where } \upsilon = [\exp(\sigma^2) - 1]^{0.5}$$

(6.79)

A lognormal random variable takes on values in the range $[0, +\infty]$. The parameter μ determines the scale of the X-distribution whereas σ^2

Fig. 6.3 Lognormal
probability density
functions with various
standard deviations σ

determines the shape of the distribution. The
mean and variance of the lognormal distribution
are given in Eq. 6.65. Figure 6.3 illustrates
the various shapes 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
$\sigma \rightarrow 0$. The density function always has a value
of zero at $x = 0$. The coefficient of variation
and skew are

$$
\begin{aligned}
\mathrm{CV}_X &= [\exp(\sigma^2) - 1]^{1/2} \\
\gamma_X &= 3\mathrm{CV}_X + \mathrm{CV}_X^3
\end{aligned}
\tag{6.80}
$$

The maximum likelihood estimates of μ and
σ^2 are given in Eq. 6.63 and the moment esti-
mates in Eq. 6.66. For reasonable-size samples,
the maximum likelihood estimates are generally
performed as well or better than the moment
estimates (Stedinger 1980).

The data in Table 6.2 were used to calculate
the parameters of the lognormal distribution that
would describe the flood flows and the results are
reported after Eq. 6.66. The two-parameter
maximum likelihood and method of moments
estimators identify parameter estimates for which
the distribution skewness coefficients 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 distri-
bution, where $X \geq \tau$. Theoretically, τ should be
positive if for physical reasons X must be posi-
tive; practically, negative values of τ can be
allowed when the resulting probability of nega-
tive values of X is sufficiently small.

Unfortunately, maximum likelihood estimates
of the parameters μ, σ^2, and τ are poorly behaved
because of irregularities in the likelihood func-
tion (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 τ by

$$
\hat{\tau} = \frac{x_{(1)}x_{(n)} - \hat{x}_{0.50}^2}{x_{(1)} + x_{(n)} - 2\hat{x}_{0.50}}
\tag{6.81}
$$

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 $x_{(1)} + x_{(n)} - 2\hat{x}_{0.50} < 0$, 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 μ and σ^2 to go with $\hat{\tau}$ in Eq. 6.81 are (Stedinger 1980)

$$\hat{\mu} = \ell n \left[\frac{\bar{x} - \hat{\tau}}{\sqrt{1 + s_x^2/(\bar{x} - \hat{\tau})^2}} \right]$$

$$\hat{\sigma}^2 = \ell n \left[1 + \frac{s_x^2}{(\bar{x} - \hat{\tau})^2} \right]$$

(6.82)

For the data in Table 6.2, Eq. 6.82 yields the hybrid moment-of-moments estimates for the three-parameter lognormal distribution

$$\hat{\mu} = 7.606$$
$$\hat{\sigma}^2 = 0.1339 = (0.3659)^2$$
$$\hat{\tau} = -600.1$$

This distribution has a coefficient of skewness of 1.19, which is more consistent with the sample skewness estimator than was the value obtained when a two-parameter lognormal distribution was fit to the data. Alternatively, one can estimate μ and σ^2 by the sample mean and variance of $\ln(X - \hat{\tau})$ that yields the hybrid maximum likelihood estimates

$$\hat{\mu} = 7.605$$
$$\hat{\sigma}^2 = 0.1407 = (0.3751)^2$$
$$\hat{\tau} = -600.1$$

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 used 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, Eq. 18.2.8).

6.3.4   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 (Bobee and Ashkar 1991). For a gamma random variable X,
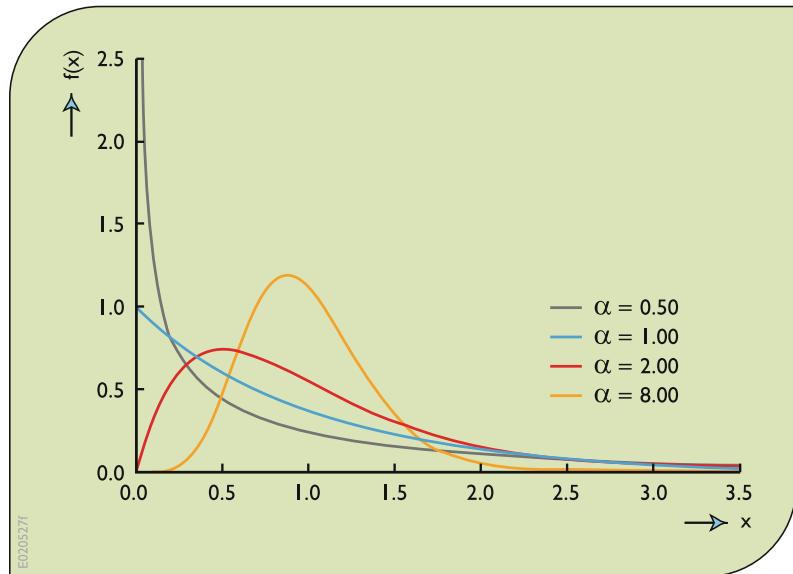
$$f_X(x) = \frac{|\beta|}{\Gamma(\alpha)} (\beta x)^{\alpha-1} e^{-\beta x} \quad \beta x \geq 0$$

$$\mu_X = \frac{\alpha}{\beta}$$

$$\sigma_X^2 = \frac{\alpha}{\beta^2}$$

$$\gamma_X = \frac{2}{\sqrt{\alpha}} = 2CV_X$$

(6.83)

The gamma function, $\Gamma(\alpha)$, for integer α is $(\alpha - 1)!$. The parameter $\alpha > 0$ determines the shape of the distribution; β is the scale parameter. Figure 6.4 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 α, though it is available in many software packages including Microsoft Excel. The gamma family includes a very special case: the exponential distribution is obtained when $\alpha = 1$.

The gamma distribution has several generalizations (Bobee and Ashkar 1991). If a constant τ is subtracted from X so that $(X - \tau)$ has a gamma distribution, the distribution of X is a three-parameter gamma distribution. This is also called a *Pearson type 3 distribution*, because the resulting distribution belongs to the third type of distributions suggested by the statistician

Fig. 6.4 The gamma
distribution functions for
various values of the shape
parameter α

Karl Pearson. Another variation is the
log-Pearson type 3 distribution obtained by fit-
ting the logarithms of X with a Pearson type 3
distribution. The log-Pearson distribution is dis-
cussed further in the next section.

The method of moments may be used to
estimate the parameters of the gamma distribu-
tion. For the three-parameter gamma distribution

$$\hat{\tau} = \bar{x} - 2\left(\frac{s_X}{\hat{\gamma}_X}\right)$$

$$\hat{\alpha} = \frac{4}{(\hat{\gamma}_X)^2} \qquad (6.84)$$

$$\hat{\beta} = \frac{2}{s_X \gamma_X}$$

where $\bar{x}, s_X^2$, and $\hat{\gamma}_X$ are estimates of the mean,
variance, and coefficient of skewness of the dis-
tribution of X (Bobee and Robitaille 1977).

For the two-parameter gamma distribution,

$$\hat{\alpha} = \frac{(\bar{x})^2}{s_x^2}$$

$$\hat{\beta} = \frac{\bar{x}}{s_x^2} \qquad (6.85)$$

Again the flood record in Table 6.2 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

$$\hat{\tau} = -735.6$$

$$\hat{\alpha} = 7.888$$

$$\hat{\beta} = 0.003452 = 1/289.7$$

Using only the sample mean and variance
yields the method of moment estimators of the
parameters of the two-parameter gamma distri-
bution ($\tau = 0$)

$$\hat{\alpha} = 3.627$$

$$\hat{\beta} = 0.002341 = 1/427.2$$

The fitted two-parameter gamma distribution
has a coefficient of skewness γ 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 esti-
mated 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 often 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 $|\gamma| \le 3$, the Wilson–Hilferty transformation

$$x_G = \mu + \sigma \left[\frac{2}{\gamma} \left(1 + \frac{\gamma x_N}{6} - \frac{\gamma^2}{36} \right)^3 - \frac{2}{\gamma} \right] \quad (6.86)$$

gives the quantiles x_G of the gamma distribution in terms of x_N, the quantiles of the standard normal distribution. Here μ, σ, and γ 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.
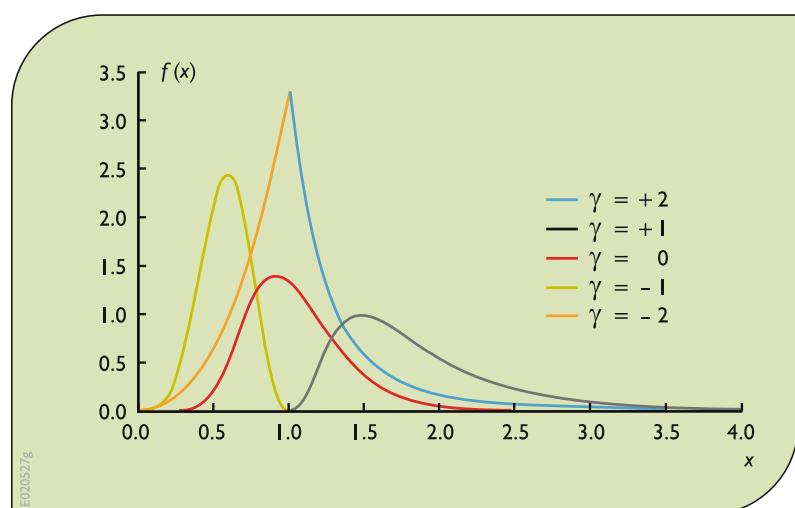
6.3.5   Log-Pearson Type 3 Distribution

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 modeling flood frequencies and has been recommended for that purpose (IACWD 1982). Bobee (1975), Bobee and Ashkar (1991) and Griffis and Stedinger (2007a) discuss the unusual shapes that this hybrid distribution may take allowing negative values of β. The LP3 distribution has a probability density function given by.

$$f_X(x) = |\beta| \{ \beta [\ln(x) - \xi] \}^{\alpha-1} \\ \exp\{-\beta[\ln(x) - \xi]\} / \{x\Gamma(\alpha)\} \quad (6.87)$$

with $\alpha > 0$, and β 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 6.5 illustrates the probability density function for the LP3

Fig. 6.5 Log-Pearson type 3 probability density functions for different values of coefficient of skewness γ

distribution as a function of the skew γ 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 2-parameter lognormal distribution is obtained as a special case.

The LP3 distribution has mean and variance

$$\mu_X = e^{\xi}\left(\frac{\beta}{\beta - 1}\right)^{\alpha}$$

$$\sigma_X^2 = e^{2\xi}\left\{\left(\frac{\beta}{\beta - 2}\right)^{\alpha} - \left(\frac{\beta}{\beta - 1}\right)^{2\alpha}\right\} \quad (6.88)$$

for $\beta > 2$, or $\beta < 0$.

For $0 < \beta < 2$, the variance is infinite.

These expressions are seldom used, but they do reveal the character of the distribution. Figures 6.6 and 6.7 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 σ_Y and coefficient of skew γ_Y of the log-transformation $Y = \ln(X)$. Thus the standard deviation σ_Y and skew γ_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 σ_Y serves as the sole shape parameter, and the coefficient of variation of X for small σ_Y is just σ_Y. Figure 6.7 shows that the situation is more complicated for the LP3 distribution. However, for small σ_Y, the coefficient of variation of X is approximately σ_Y.

Again, the flood flow data in Table 6.2 can be used to illustrate parameter estimation. Using natural logarithms, one can estimate the log-space moments with the standard estimators in Eqs. 6.39a that yield
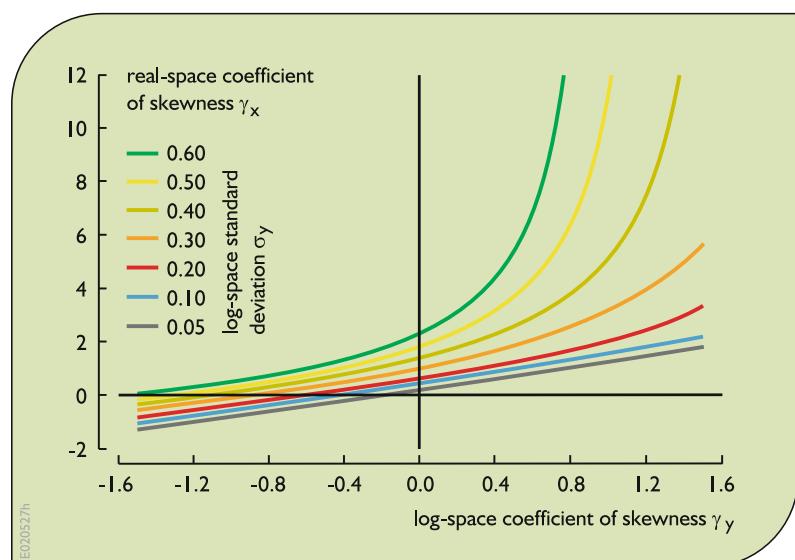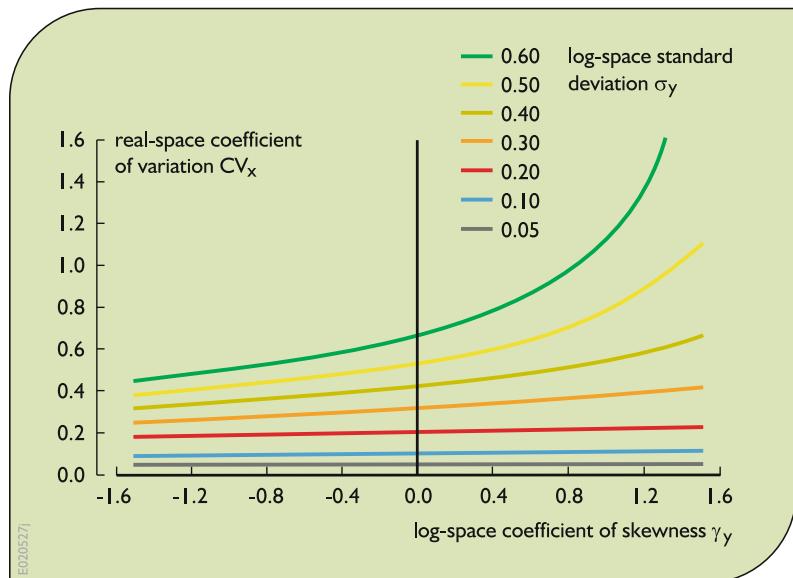
$$\hat{\mu} = 7.202$$
$$\hat{\sigma} = 0.5625$$
$$\hat{\gamma} = -0.337$$

For the LP3 distribution, analysis generally focuses on the distribution of the logarithms $Y = \ln(X)$ of the flows, which would have a

Fig. 6.6 Real-space coefficient of skewness γ_X for LP3 distributed X as a function of log-space standard deviation σ_Y and coefficient of skewness γ_Y where $Y = \ln(X)$

Fig. 6.7 Real-space
coefficient of variation CV_X
for LP3 distributed X as a
function of log-space
standard deviation σ_Y and
coefficient of skewness γ_Y
where $Y = \ln(X)$

Pearson type 3 distribution with moments μ_Y, σ_Y and γ_Y (IACWD 1982; Bobée and Ashkar 1991). As a result, flood quantiles are calculated as

$$x_p = \exp\{\mu_Y + \sigma_Y K_p[\gamma_Y]\} \qquad (6.89)$$

where $K_p[\gamma_Y]$ is a frequency factor corresponding to cumulative probability for skewness coefficient γ_Y. ($K_p[\gamma_Y]$ corresponds to the quantiles of a three-parameter gamma distribution with zero mean, unit variance, and skewness coefficient γ_Y.)

Since 1967, the recommended procedure for flood frequency analysis by federal agencies in the United States uses this distribution. Current guidelines in Bulletin 17B (IACWD 1982) suggest that the skew γ_Y be estimated by a weighted average of the at-site sample skewness coefficient and a regional estimate of the skewness coefficient. Griffis and Stedinger (2007b) compare a wide range of methods that have been recommended for fitting the LP3 distribution.

6.3.6 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

$$F_X(x) = \exp\{-\exp[-(x - \xi)/\alpha]\} \qquad (6.90)$$

with mean and variance of

$$\begin{aligned} \mu_X &= \xi + 0.5772\alpha \\ \sigma_X^2 &= \pi^2\alpha^2/6 \cong 1.645\alpha^2 \end{aligned} \qquad (6.91)$$

Its skewness coefficient has the 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 cumulative distribution function

$$F_X(x) = \exp\{-[1 - \kappa(x - \xi)/\alpha]^{1/\kappa}\} \quad \text{for } \kappa \neq 0 \tag{6.92}$$

For $\kappa > 0$, floods must be less than the upper bound for $\kappa < 0$, $\xi < x < \infty$, whereas for $\kappa > 0$, $\xi < x < \xi + \alpha/\kappa$ (Hosking and Wallis 1987). The mean, variance, and skewness coefficient are (for $\kappa > -1/3$)

$$\mu_X = \xi + (\alpha/\kappa)[1 - \Gamma(1+\kappa)],$$
$$\sigma_X^2 = (\alpha/\kappa)^2\{\Gamma(1+2\kappa) - [\Gamma(1+\kappa)]^2\}$$
$$\gamma_X = \text{Sign}(\kappa)\{-\Gamma(1+3\kappa) + 3\Gamma(1+\kappa)\Gamma(1+2\kappa)$$
$$- 2[\Gamma(1+\kappa)]^3\}/\{\Gamma(1+2\kappa) - [\Gamma(1+\kappa)]^2\}^{3/2} \tag{6.93}$$

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 Figs. 6.8 and 6.9.

The parameters of the GEV distribution are easily computed using L-moments and the relationships (Hosking et al. (1985)

$$\kappa = 7.8590c + 2.9554c^2$$
$$\alpha = \kappa\lambda_2/[\Gamma(1+\kappa)(1 - 2^{-\kappa})] \tag{6.94}$$
$$\xi = \lambda_1 + (\alpha/\kappa)[\Gamma(1+\kappa) - 1]$$

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 κ will depend only upon the L-skewness estimator $\hat{\tau}_3$. The estimator of the scale parameter α will then depend on the estimate of κ and of λ_2. Finally, one must also use the sample mean λ_1 (Eq. 6.48) to determine the estimate of the location parameter ξ.
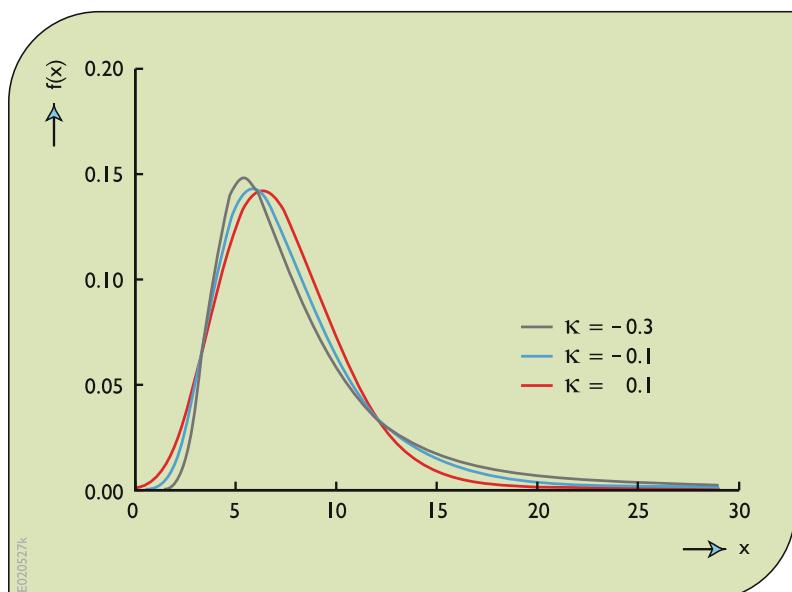
Using the flood data in Table 6.2 and the sample L-moments computed in Sect. 6.2, one obtains first

$$c = -0.000896$$
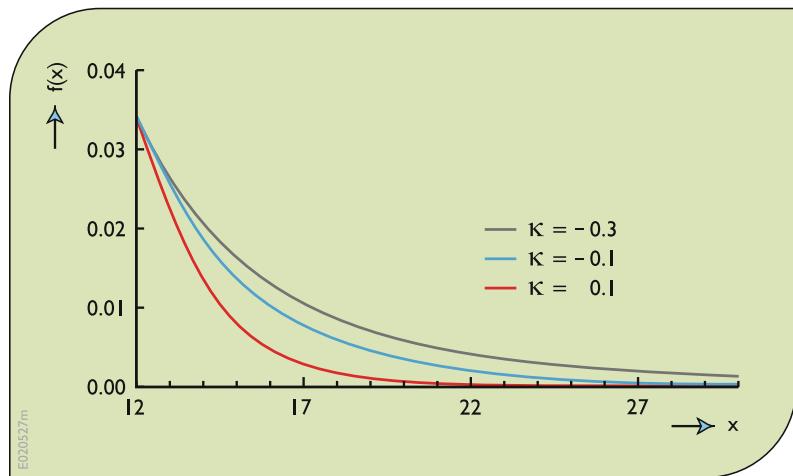
that yields

$$\hat{\kappa} = -0.007036$$
$$\hat{\xi} = 1165.20$$
$$\hat{\alpha} = 657.29$$

Fig. 6.8 GEV density distributions for selected shape parameter κ values

Fig. 6.9 Right-hand tails
of GEV distributions
shown in Fig. 6.8

The small value of the fitted κ parameter means that the fitted distribution is essentially a Gumbel distribution. Here ξ 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. Yet Martins and Stedinger (2001a, b) found that with occasional uninformative samples, the MLE estimator of κ could be entirely unrealistic resulting in absurd quantile estimators. However the use of a realistic prior distribution on κ yielded better generalized maximum likelihood estimators (GLME) than moment and L-moment estimators over the range of κ of interest.

The GMLE estimators are obtained my maximizing the log-likelihood function, augmented by a prior density function on κ. A prior distribution that reflects general worldwide geophysical experience and physical realism is in the form of a beta distribution

$$\pi(\kappa) = \Gamma(p)\Gamma(q)(0.5 + \kappa)^{p-1}$$
$$(0.5 - \kappa)^{q-1}/\Gamma(p+q) \qquad (6.95)$$

for $-0.5 < \kappa < +0.5$ with $p = 6$ and $q = 9$. Moreover, this prior assigns reasonable probabilities to the values of κ within that range. For κ 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 κ 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 $\{x_1, \ldots, x_n\}$ drawn for a GEV distribution, the generalized likelihood function is

$$\ln\{L(\xi, \alpha, \kappa | x_1, \ldots, x_n)\}$$
$$= -n\ln(\alpha) + \sum_{i=1}^{n}\left[\left(\frac{1}{\kappa} - 1\right)\ln(y_i) - (y_i)^{1/\kappa}\right]$$
$$+ \ln[\pi(\kappa)]$$

with

$$y_i = [1 - (\kappa/\alpha)(x_i - \xi)] \qquad (6.96)$$

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 $\min\{y_1, \ldots, y_n\} \geq \varepsilon$ 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 6.2 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 Eq. 6.96 yielded the maximum likelihood estimators of the GEV parameters

$$\hat{\kappa} = -0.0359$$
$$\hat{\xi} = 1165.4$$
$$\hat{\alpha} = 620.2$$

Similarly, use of the geophysical prior (Eq. 6.95) yielded the generalized maximum likelihood estimators

$$\hat{\kappa} = -0.0823$$
$$\hat{\xi} = 1150.8$$
$$\hat{\alpha} = 611.4$$

Here the record length of 40 years is too short to reliably define the shape parameter κ so that result of using the prior is to increase κ slightly toward the mean of the prior. The other two parameters adjust accordingly.

6.3.7   L-Moment Diagrams
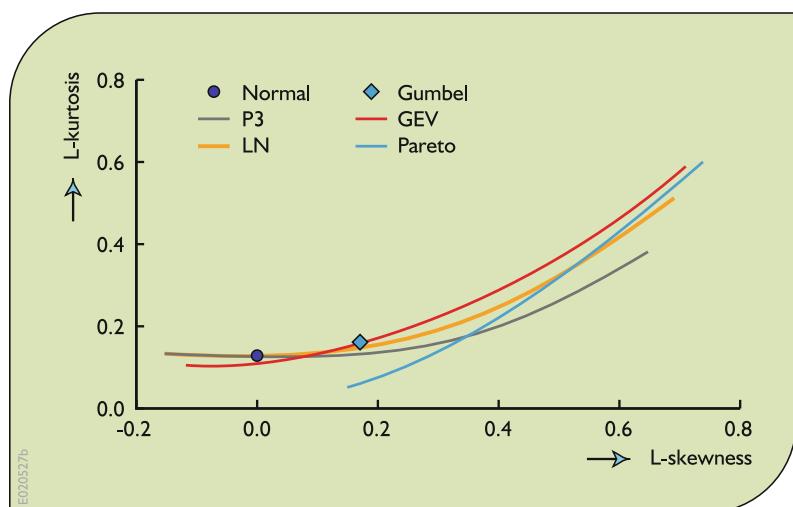
This chapter has presented several families of distributions. The L-moment diagram in Fig. 6.10 illustrates the relationships between the L-kurtosis (τ_3) and L-skewness (τ_2) for a number of the families of distributions often used in

hydrology. It shows that distributions with the same coefficient of skewness still differ in the thickness of their tails, 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 κ yielding almost symmetric or even negatively skewed GEV distributions, the GEV has a smaller kurtosis than the three-parameter lognormal distribution. The latter can

Fig. 6.10 Relationships between L-skewness and L-kurtosis for various distributions

be negatively skewed when τ is used as an upper bound.

Figurer 6.10 also includes the three-parameter generalized Pareto distribution, whose cdf is

$$F_X(x) = 1 - [1 - \kappa(x - \xi)/\alpha]^{1/\kappa} \qquad (6.97)$$

(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 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 α 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 Fig. 6.10, the GEV distribution has a thicker right-hand tail than either the gamma/Pearson type 3 distribution or the lognormal distribution.

6.4   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 based on the memories of humans who have experienced these events and/or physical markings in the watershed (Stedinger and Baker 1987). Before continuous measurements were taken that provided this information, the annual maximum floods that were not unusual were not recorded. These missing data are censored data. They cover periods between occasionally 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 with censored data 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 "nonzero" values of *X*. Then the parameters of the *G* distribution can be estimated using any procedure appropriate for complete uncensored samples. The unconditional cumulative distribution function (cdf) $F_X(x)$ for any value $x > 0$, is then

$$F_X(x) = P_0 + (1 - P_0)G(x) \qquad (6.98)$$

This model completely decouples the value of P_0 from the parameters that describe the *G* distribution.

Section 6.3.2 discusses probability plots and plotting positions useful for graphical displaying 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 $i = 1, …, r$, wherein $x_{(r)}$ is the largest, then reasonable plotting positions within the interval $[P_0, 1]$ are

$$p_i = P_0 + (1 - P_0)[(i - a)/(r + 1 - 2a)] \qquad (6.99)$$

where *a* defines the plotting position that is used. Helsel and Cohn (1988) show that reasonable choices for *a* generally make little difference. Letting $a = 0$ is reasonable (Hirsch and Stedinger 1987). 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 $r \gg (1 - 2a)$, p_i is indistinguishable from

$$p_i' = [i + (n + r) - a]/(n + 1 - 2a). \qquad (6.100)$$

where again, $i = 1, …, r$. 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, of the standard deviation, and of the 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 $i = 1, …, r$, where the *r* largest observation in a sample of size *n* are available. If regression yields constant *m* and slope *s*, corresponding to population moments μ and σ, a good estimator of the *p*th quantile is

$$x_p = \exp[m + s z_p] \qquad (6.101)$$

where z_p is the *p*th quantile of the standard normal distribution. To estimate sample means and other statistics one can fill in the missing observations with

$$x(j) = \exp\{y(j)\} \quad \text{for } j = 1, …, (n - r) \qquad (6.102)$$

where

$$y(j) = m + s F^{-1}\{P_0[(j - a)/(n - r + 1 - 2a)]\} \qquad (6.103)$$

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 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, expected moment estimators seem to do as well as MLEs with the LP3 distribution (Cohn et al. 1997, 2001).

While often a computational challenge, maximum likelihood estimators for complete samples, and samples with some observations censored, pose no conceptual challenge. One need to only write the maximum likelihood function for the data and proceed to seek the parameter values that maximizes that function. Thus if $F(x|\theta)$ and $f(x|\theta)$ are the cumulative distribution and probability density functions that should describe the data, and θ are its parameters, 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)

$$L(\theta|r, n, x_1, \ldots, x_r) = F(T|\theta)^{(n-r)} f(x_1|\theta) f(x_2|\theta) \ldots f(x_r|\theta)$$

$$(6.104)$$

Here $(n - r)$ observations were below the threshold T, and the probability an observation is below T is $F(T|\theta)$ which then appears in Eq. 6.104 to represent that observation. In addition the specific values of the r observations $x_1, \ldots, x_r$ are available. The probability an observation is in a small interval of width δ^r around x_i is $\delta^r f(x_i|\theta)$. Thus strictly speaking the likelihood function also includes a term δ^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 $[1 - F(M|\theta)]$ in the likelihood function. Similarly, if all that was known was that the value was between L and M, then a term $[F(M|\theta) - F(L|\theta)]$ 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.

6.5 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" 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 pth quantile x_p is then reduced to estimating the mean for a site μ_x, and the ratio x_p/μ_x of the pth 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 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 κ 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 weighting 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 κ-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 semiarid 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 size for different parameters. For realistic hydrologic regions, these and other studies illustrate the value of regionalizing estimators of the shape, and often the coefficient of variation of a distribution.

6.6   **Partial Duration Series**

Two general approaches are available for modeling 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 semiarid 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 modeling all events that occur within a year that exceed some threshold of interest.

Use of a partial duration series (PDS) framework avoids such problems by considering all independent peaks that exceed a specified threshold. And, one can estimate annual exceedance probabilities from the analysis of PDS. Arguments in favor of PDS are that relatively long and reliable PDS 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), PDS 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 using a PDS approach over an annual maximum approach, 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 PDS analyses is that one must have criteria to identify only independent peaks (and not multiple peaks corresponding to the same event). Thus PDS 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 hydrologic events when more than one damage-causing event can occur in a season or within a year (North 1980).

A model of a PDS 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 PDS model, let λ 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

$$1/T_a = 1 - \exp\{-\lambda q_e\} = 1 - \exp\{-1/T_p\}$$
$$(6.105)$$

where $T_p = 1/(\lambda q_e)$ is the average return period for level x in the PDS.

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

$$F_X(x) = 1 - [1 - \kappa(x - \xi)/\alpha]^{1/\kappa} \qquad (6.106)$$

with mean and variance

$$\mu_X = \xi + \alpha/(1 + \kappa)\kappa$$
$$\sigma_X^2 = \alpha^2/[(1 + \kappa)^2(1 + 2\kappa)] \qquad (6.107)$$

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 obtain 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 are a very reasonable description 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 κ from annual maximum and PDS analyses can be used interchangeably. Martins and Stedinger (2001a, b) compare PDS estimation procedures a well as demonstrating that use of the three-parameter Poisson-GPD model instead of a three-parameter GEV distribution generally results in flood quantile estimators with the same precision.

Madsen and Rosbjerg (1997a) use a Poisson-GPD model as the basis of a PDS 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 κ 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.

6.7 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 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 generation and optimization models. The third part discusses the sampling properties of statistics used to describe the characteristics of many time series.

6.7.1 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 index $t = 1, 2, 3, \ldots$.

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, at least in the short run. This means that the probability distribution of the process is not changing over some specified interval of 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. In other words, its shape does not change over time if the distribution is stationary. In the long run, however, because of climate and land changes, many hydrologic distributions are not stationary, and

just how much they will change in the future is uncertain.

For a stationary stochastic process, one can write the mean and variance as

$$\mu_X = E[X(t)] \qquad (6.109)$$

and

$$\sigma_X^2 = Var[X(t)] \qquad (6.110)$$

Both are independent of time t. The *autocorrelations*, the correlation of X with itself, are given by

$$\rho_X(k) = \frac{Cov[X(t), X(t+k)]}{\sigma_X^2} \qquad (6.111)$$

for any positive integer k (the time lag). 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 μ_X, σ_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 Eq. 6.39a.

$$\hat{\mu}_X = \overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t \qquad (6.112)$$

$$\hat{\sigma}_X^2 = \frac{1}{T}\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2 \qquad (6.113)$$

while the autocorrelations $\rho_X(k)$ can be estimated as (Jenkins and Watts 1968)

$$\hat{\rho}_x(k) = r_k = \frac{\sum_{t=1}^{T-k}(x_{t+k} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \qquad (6.114)$$

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 6.7.3 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 Eqs. 6.109–6.111 have the intended meaning. Stochastic processes are not always stationary. Urban development, deforestation, agricultural development, 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 cannot be employed and the problem generally becomes much more difficult.

6.7.2 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. In symbols for $k > 0$,

$$F_X[X(t+k)|X(t), X(t-1), X(t-2), \ldots] = F_X[X(t+k)|X(t)]$$

$$(6.115)$$

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 processes 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 simpler 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

$$\sum_{i=1}^{n} p_i = 1 \qquad (6.116)$$

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},

$$p_{ij} = \Pr[Q_{y+1} = q_j | Q_y = q_i] \qquad (6.117)$$

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

$$\sum_{j=1}^{n} p_{ij} = 1 \quad \text{for all } i \qquad (6.118)$$

Figure 6.11a, b show a possible set of transition probabilities in a matrix and as histograms. 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.*

Let 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 behavior of the process. Let p_i^y be the

probability that the process resides in state i in year y. Then the probability that $Q_{y+1} = q_j$ is the sum of the probabilities p_i^y that $Q_y = q_i$ times the probability p_{ij} that the next state is Q_{y+1} given that $Q_y = q_i$. In symbols, this relationship is written

$$p_j^{y+1} = p_1^y p_{1j} + p_2^y p_{2j} + \cdots + p_n^y p_{nj} = \sum_{i=1}^{n} p_i^y p_{ij}$$
$$(6.119)$$

Letting p^y be the row vector of state resident probabilities $(p_i^y, \ldots, p_n^y)$, this relationship may be written

$$p^{(y+1)} = p^{(y)} P \qquad (6.120)$$
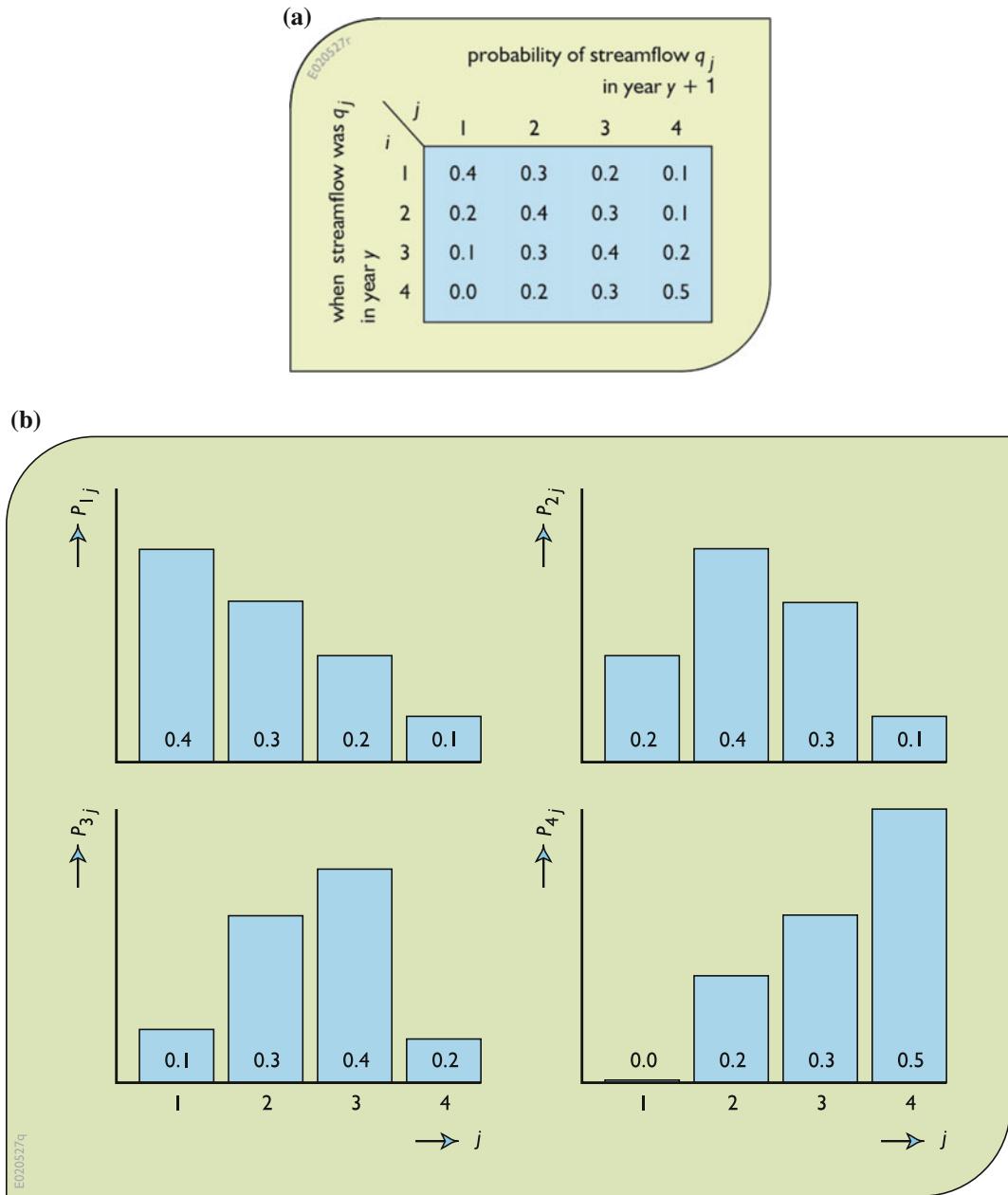
To calculate the probabilities of each streamflow state in year $y + 2$, one can use $p^{(y+1)}$ in Eq. 6.120 to obtain $p^{(y+2)} = p^{(y+1)} P$ or $p^{(y+2)} = p^y P^2$

Continuing in this matter, it is possible to compute the probabilities of each possible streamflow state for years $y + 1$, $y + 2$, $y + 3$, …, $y + k$, … as

$$p^{(y+k)} = p^y P^k \qquad (6.121)$$

Returning to the four-state example in Fig. 6.11, 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 Fig. 6.11, 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 6.8.

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 $p^{(y+k)}$ approaches for large k. It is clear from Table 6.8 that as k becomes larger, Eq. 6.119 becomes

Fig. 6.11   a Matrix 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. **b** 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

Table 6.8 Successive streamflow probabilities based on transition probabilities in Fig. 6.11

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

$$p_j = \sum_{i=1}^{n} p_i p_{ij} \qquad (6.122)$$

or in vector notation, Eq. 6.122 becomes

$$p = pP \qquad (6.123)$$

where p is the row vector of unconditional probabilities $(p_1, \ldots, p_n)$. For the example in Table 6.8, the probability vector p equals (0.156, 0.309, 0.316, 0.219).

The steady-state probabilities for any Markov chain can be found by solving simultaneous Eqs. 6.123 for all but one of the states j together with the constraint

$$\sum_{i=1}^{n} p_i = 1 \qquad (6.124)$$

Annual streamflows are seldom as highly correlated as the flows in this example. However, monthly, weekly, and especially daily stream-flows generally have high serial correlations. Assuming that the unconditional steady-state probability distributions for monthly stream-flows are stationary, a Markov chain can be defined for each month's steamflow. Since there

are 12 months in a year, there would be 12 transition matrices, the elements of which could be denoted as p_{ij}^t. Each defines the probability of a streamflow $p_j^{t+1}(y)$ in month $t + 1$, given a streamflow $p_i^t(y)$ in month t. The steady-state stationary probability vectors for each month can be found by the procedure outlined above, except that now all 12 matrices are used to calculate all 12 steady-state probability vectors. However, once the steady-state vector p is found for one month, the others are easily computed using Eq. 6.121 with t replacing y.

6.7.3 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 the case in time series, can have a marked effect on the distribution of these statistics. This part of Sect. 6.7 reviews the sampling properties of these statistics when the observations are a realization of a stochastic process.

The sample mean

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad (6.125)$$

when viewed as a random variable is an unbiased estimate of the mean of the process μ_X, because

$$E[\overline{X}] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \mu_X \qquad (6.126)$$

However, correlation among the X_i's, so that $\rho_X(k) \neq 0$ for $k > 0$, affects the variance of the estimated mean $\overline{X}$.

$$\begin{aligned}
\mathrm{Var}(\overline{X}) &= E\left[(\overline{X} - \mu_X)^2\right] \\
&= \frac{1}{n^2} E\left\{ \sum_{t=1}^{n} \sum_{s=1}^{n} (X_t - \mu_X)(X_s - \mu_X) \right\} \\
&= \frac{\sigma_X^2}{n} \left\{ 1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \rho_X(k) \right\}
\end{aligned}$$
$$(6.127)$$

The variance of $\overline{X}$, equal to σ_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 non-decreasing function of n, so that the variance of $\overline{X}$ is inflated by a factor whose importance does not decrease with increasing sample size. This is an important observation, because it means 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

$$\rho_X(k) = [\rho_X(1)]^k = \rho^k \qquad (6.128)$$

This correlation structure arises from the autoregressive Markov model discussed at length in Sect. 6.8. For this correlation structure

$$\mathrm{Var}(\overline{X}) = \frac{\sigma_X^2}{n} \left\{ 1 + \frac{2\rho}{n} \frac{[n(1 - \rho) - (1 - \rho^n)]}{(1 - \rho)^2} \right\}$$
$$(6.129)$$

Substitution of the sample estimates for σ_X^2 and $\rho_X(1)$ in the equation above often yields a more realistic estimate of the variance of $\overline{X}$ than does the estimate s_X^2/n if the correlation structure $\rho_X(k) = \rho^k$ is reasonable; otherwise, Eq. 6.127 may be employed. Table 6.9 illustrates the affect of correlation among the X_t values on the standard error of their mean, equal to the square root of the variance in Eq. 6.127.

The properties of the estimate of the variance of X,

$$\hat{\sigma}_X^2 = v_X^2 = \frac{1}{n} \sum_{t=1}^{n} (X_t - \overline{X})^2 \qquad (6.130)$$

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

Table 6.9 Standard error of $\overline{X}$ when $\sigma_x = 0.25$ and $\rho_X(k) = \rho^k$

sample size n	correlation of consecutive observations		
	p = 0.0	p = 0.3	p = 0.6
25	0.050	0.067	0.096
50	0.035	0.048	0.069
100	0.025	0.034	0.050

E02I10Ij

denominator rather than $n - 1$. The expected value of v_x^2 becomes

$$E[v_X^2] = \sigma_X^2 \left\{ 1 - \frac{1}{n} - \frac{2}{n} \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \rho_X(k) \right\}$$
(6.131)

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, Sect. 48.1).

$$\text{Var}(v_X^2) \cong 2 \frac{\sigma_X^4}{n} \left\{ 1 + 2 \sum_{k=1}^{\infty} \rho_X^2(k) \right\}$$
(6.132)

where for $\rho_X(k) = \rho^k$, Eq. 6.132 becomes

$$\text{Var}(v_X^2) \cong 2 \frac{\sigma_X^4}{n} \left(\frac{1 + \rho^2}{1 - \rho^2} \right)$$
(6.133)

Like the variance of $\overline{X}$, the variance of v_X^2 is inflated by a factor whose importance does not decrease with n. This is illustrated by Table 6.10 that gives the standard deviation of v_X^2 divided by the true variance σ_X^2 as a function of n and ρ 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 between the random variable 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

$$\hat{\rho}_X(k) = r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2}$$
(6.134)

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)

$$E[r_1] = -\frac{1}{n}$$
(6.135a)

and

$$\text{Var}(r_1) = \frac{(n-2)^2}{n^2(n-1)} \cong \frac{1}{n}$$
(6.135b)

Table 6.10 Standard deviation of (v_X^2/σ_X^2) when observations have a normal distribution and $\rho_X(k) = \rho^k$

sample size n	correlation of consecutive observations		
	$p = 0.0$	$p = 0.3$	$p = 0.6$
25	0.28	0.31	0.41
50	0.20	0.22	0.29
100	0.14	0.15	0.21

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 $(E[r_1])^2 + \text{Var}(r_1) = 0.0016 + 0.0353 = 0.0369$. 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. However, for normally distributed X_t and large n (Kendall and Stuart 1966),

$$\text{Var}(r_k) \cong \frac{1}{n} \sum_{l=-\infty}^{+\infty} \left[\rho_x^2(l) + \rho_x(l+k)\rho_x(l-k) \right.$$
$$\left. - 4\rho_x(k)\rho_x(l)\rho_x(k-l) + 2\rho_x^2(k)\rho_x^2(l) \right]$$
$$(6.136)$$

If $\rho_X(k)$ is essentially zero for $k > q$, then the simpler expression (Box et al. 1994)

$$\text{Var}(r_k) \cong \frac{1}{n} \left[1 + 2 \sum_{t=1}^{q} \rho_x^2(l) \right] \qquad (6.137)$$

is valid for r_k corresponding to $k > q$; thus for large n, $\text{Var}(r_k) \geq l/n$ 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 $\rho_X(k) = \rho^k$, Eq. 6.137 reduces to

$$\text{Var}(r_k) \cong \frac{1}{n} \left[\frac{(1+\rho^2)(1-\rho^{2k})}{1-\rho^2} - 2k\rho^{2k} \right]$$
$$(6.138)$$

In particular for r_1, this gives

$$\text{Var}(r_1) \cong \frac{1}{n} \left(1 - \rho^2 \right) \qquad (6.139)$$

Approximate values of the standard deviation of r_1 for different values of n and ρ are given in Table 6.11.

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.

6.8 Synthetic Streamflow Generation

6.8.1 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

Table 6.11 Approximate standard deviation of r_1 when observations have a normal distribution and $\rho_X(k) = \rho^k$

sample size n	correlation of consecutive observations		
	$p = 0.0$	$p = 0.3$	$p = 0.6$
25	0.20	0.19	0.16
50	0.14	0.13	0.11
100	0.10	0.095	0.080

E02II0Im

because of the historic 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 field has been called stochastic hydrologic modeling. More detailed presentations can be found in Marco et al. (1989) and Salas (1993).
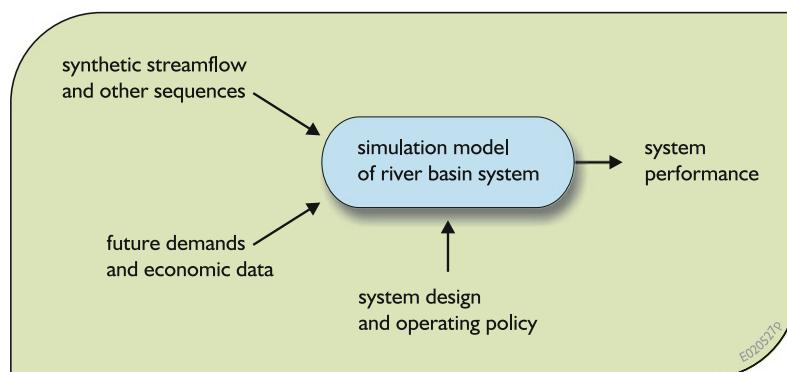
River basin simulation studies can use many sets of streamflow, rainfall, evaporation, and/or temperature sequences to evaluate the statistical properties of the performance of alternative water resources systems. For this purpose, synthetic flows and other generated quantities should resemble, statistically, those sequences that are likely to be experienced during the planning period. Figure 6.12 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. By 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 likely, 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 resource systems having large amounts of over-year storage. Use of only the historical hydrologic record in system simulation yields only one time history of how the system would operate from year to year. In water resource systems having relatively little storage so that reservoirs and/or groundwater aquifers refill almost every year, synthetic hydrologic sequences may not be needed if historical sequences of a reasonable length are available. In this second case, a 25-year historic record provides 25 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 resource system performance indices can

Fig. 6.12 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

be estimated, and some studies have shown this to be the case (Vogel and Shallcross 1996; Vogel and Stedinger 1988). In particular, if the operation of the system and performance indices have thresholds and shape breaks, then the coarse description 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 a nonlinear penalty function on average one year in 20. Then in a 60-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 now 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 sufficient 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: 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 to just 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 nonlinear, then modeling 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 that it deals very well with the few droughts in the historical record but at the same time 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 Pei and Cohn 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 that 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 will 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.

6.8.2 Streamflow Generation Models

A statistical streamflow generation model is used to generate streamflow data that can supplement or replace historical streamflow data in various analyses requiring such data. If the past flow record is considered representative of what the future one might be, at least for a while, then the statistical characteristics of the historical flow record can be used as a basis for generating new flow data. While this may be a reasonable assumption in the near future, changing land uses and climate may lead to entirely different statistical characteristics of future streamflows, if not now, certainly in the more distant future. By then, improved global climate models (GCMs) and downscaling methods together with improved rainfall-runoff predictions given future land use scenarios may be a preferred way to generate future streamflows. This section of the chapter will focus on the use of historical records.

The first step in the construction of a statistical streamflow generating model based on historical flow records 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, modeling 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, b; Thyer and Kuczera 2000). For multicomponent 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 statistically resemble the historical flows if the streamflow model produces flows with the same mean, variance, skew coefficient, autocorrelations, and/or cross-correlations as were observed in the historic series. This definition of statistical resemblance is attractive because it is operational and requires that an analyst need only find a model that can reproduce the observed statistics. The drawback of this approach is that it shifts the modeling 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 Sect. 6.7) 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 modeled. The decision should depend on what characteristics are important to the operation of the system being studied, the data available, 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 using an alternative model and set of parameter values. If the model's results are sensitive to changes, then, as always, one must exercise judgment 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 data. Section 6.8.3 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. Therefore, considerable time is spent 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 to generate concurrent flows at several sites and to generate seasonal or monthly flows while preserving the characteristics of annual flows. For those wishing to study synthetic streamflow models in greater depth more advanced material can be found in Marco et al. (1989) and Salas (1993).

6.8.3 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 a particular value 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 different periods. A common and reasonable assumption is that annual flows are the result of a first-order Markov process.

Assume also that annual streamflows are normally distributed. In some areas, the distribution of annual flows is in fact nearly normal. Streamflow models that produce nonnormal 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 μ, variance σ^2, and year-to-year correlation ρ between flows is

$$f(q_y, q_w) = \frac{1}{2\pi\sigma^2(1-\rho^2)^{0.5}}$$
$$\cdot \exp\left[\frac{(q_y-\mu)^2 - 2\rho(q_y-\mu)(q_w-\mu) + (q_w-\mu)^2}{2\sigma^2(1-\rho^2)}\right]$$

$$(6.140)$$

The joint normal distribution for two random variables with the same mean and variance depend only on their common mean μ, variance σ^2, and the correlation ρ 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 dependence of the distribution of the flow in year $y + 1$ on flows in previous years 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

$$E[Q_{y+1}|Q_y = q_y] = \mu + \rho(q_y - \mu)$$
$$\text{Var}(Q_{y+1}|Q_y = q_y) = \sigma^2(1 - \rho^2)$$

$$(6.141)$$

where q_y is the value of Q_y in year y. Notice that the larger the absolute value of the correlation ρ 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.

Synthetic normally distributed streamflows that have mean μ, variance σ^2, and year-to-year correlation ρ, are produced by the model

$$Q_{y+1} = \mu + \rho(Q_y - \mu) + V_y\sigma\sqrt{1 - \rho^2} \tag{6.142}$$

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

$$E\left[V_w V_y\right] = 0 \quad w \neq y \tag{6.143}$$

and

$$E[(Q_w - \mu)V_y] = 0 \quad w \leq y \tag{6.144}$$

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

$$E[Q_{y+1}|q_y] = E[\mu + \rho(q_y - \mu) + V_y\sigma\sqrt{1 - \rho^2}]$$
$$= \mu + \rho(q_y - \mu) \tag{6.145}$$

Since $E\{V_y^2\} = \text{Var}[V_y] = 1$, the conditional variance of Q_{y+1} is

$$\begin{aligned} \text{Var}[Q_{y+1}|q_y] &= E[\{Q_{y+1} - E[Q_{y+1}|q_y]\}^2|q_y] \\ &= E[\{\mu + \rho(q_y - \mu) + V_y\sigma\sqrt{1 - \rho^2} \\ &\quad - [\mu + \rho(q_y - \mu)]\}^2 \\ &= E[V_y\sigma\sqrt{1 - \rho^2}]^2 = \sigma^2(1 - \rho^2)] \end{aligned} \tag{6.146}$$

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 Eq. 6.142 to obtain

$$\begin{aligned} E[Q_{y+1}] &= \mu + \rho(E[Q_y] - \mu) \\ &\quad + E[V_y]\sigma\sqrt{1 - \rho^2} \end{aligned} \tag{6.147}$$

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

$$E[Q] = \mu \tag{6.148}$$

Alternatively, if Q_y for $y = 1$ has mean μ, then Eq. 6.147 indicates that Q_2 will have mean μ. Thus repeated application of the Eq. 6.147 would demonstrate that all Q_y for $y > 1$ have mean μ.

The unconditional variance of the annual flows can be derived by squaring both sides of 6.142 to obtain

$$\begin{aligned} E[(Q_{y+1} - \mu)^2] &= E[\{\rho(Q_y - \mu) + V_y\sigma\sqrt{1 - \rho^2}\}^2] \\ &= \rho^2 E[(Q_y - \mu)^2] + 2\rho\sigma\sqrt{1 - \rho^2}E[(Q_y - \mu)V_y] \\ &\quad + \sigma^2(1 - \rho^2)E[V_y^2] \end{aligned} \tag{6.149}$$

Because V_y is independent of Q_y (Eq. 6.144), the second term on the right-hand side of Eq. 6.149 vanishes. Hence the unconditional variance of Q satisfies

$$E[(Q_{y+1} - \mu)^2] = \rho^2 E[(Q_y - \mu)^2] + \sigma^2(1 - \rho^2) \tag{6.150}$$

Assuming that Q_{y+1} and Q_y have the same variance yields

$$E[(Q - \mu)^2] = \sigma^2 \qquad (6.151)$$

so that the unconditional variance is σ^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 σ^2, then Q_y for $y \geq 1$ has variance σ^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 Eq. 6.142 one can compute that the covariance of consecutive flows must be.

$$
\begin{aligned}
E[(Q_{y+1} - \mu)(Q_y - \mu)] &= E\{[\rho(Q_y - \mu) \\
&\quad + V_y \sigma \sqrt{1 - \rho^2}](Q_y - \mu)\} \\
&= \rho E[(Q_y - \mu)^2] = \rho \sigma^2
\end{aligned}
$$
$$(6.152)$$

where $E[(Q_y - \mu)V_y] = 0$ because V_y and Q_y are independent (Eq. 6.144).

Over a longer time scale, another property of this model is that the covariance of flows in year y and $y + k$ is

$$E[(Q_{y+k} - \mu)(Q_y - \mu)] = \rho^k \sigma^2 \qquad (6.153)$$

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

$$
\begin{aligned}
E[(Q_{y+j} - \mu)(Q_y - \mu)] &= E\{[\rho(Q_{y+j-1} - \mu) \\
&\quad + V_{y+j-1} \sigma \sqrt{1 - \rho^2}](Q_y - \mu)\} \\
&= \rho E[(Q_y - \mu)(Q_{y+j-1} - \mu)] \\
&= \rho[\rho^{j-1} \sigma^2] = \rho^j \sigma^2
\end{aligned}
$$
$$(6.154)$$

where $E[(Q_y - \mu)V_{y+j-1}] = 0$ for $j \geq 1$. Hence Eq. 6.153 is true for any value of k.

It is important to note that the results in Eqs. 6.145 to 6.153 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 Eq. 6.142 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 μ and variance σ^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 μ and variance σ^2. The next section considers how this and other models can be used to generate streamflows that have other than a normal distribution.

6.8.4 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 hydrologic data cannot really be normally distributed because of the impossibility of negative values. In general, distributions of hydrologic 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 Figs. 6.3 and 6.4.

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

$$
\begin{aligned}
E[(Q_{y+1} - \mu)^3] &= E\left[\rho(Q_y - \mu) + V_y \sigma \sqrt{1 - \rho^2}\right]^3 \\
&= \rho^3 E[(Q_y - \mu)^3] \\
&\quad + \sigma^3 (1 - \rho^2)^{3/2} E\left[V_y^3\right]
\end{aligned}
$$
$$(6.155)$$

so that

$$\gamma_Q = \frac{E[(Q - \mu)^3]}{\sigma^3} = \frac{(1 - \rho^2)^{3/2}}{1 - \rho^3} \qquad (6.156)$$

By appropriate choice of the skew of V_y, 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 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

$$Q_y = \tau + \exp(X_y) \qquad (6.157)$$

where X_y is a normal random variable; when the lower bound τ is zero, Q_y has a two-parameter lognormal distribution. Equation 6.157 transforms the normal variates X_y into lognormally distributed streamflows. The transformation is easily inverted to obtain

$$X_y = \ln(Q_y - \tau) \quad \text{for } Q_y \geq \tau \qquad (6.158)$$

where Q_y must be greater than its lower bound τ.

The mean, variance, skewness of X_y and Q_y are related by the formulas (Matalas 1967)

$$\begin{aligned}
\mu_Q &= \tau + \exp(\mu_X + \tfrac{1}{2}\sigma_X^2) \\
\sigma_Q^2 &= \exp(2\mu_X + \sigma_X^2)[\exp(\sigma_X^2) - 1] \\
\gamma_Q &= \frac{\exp(3\sigma_X^2) - 3\exp(\sigma_X^2) + 2}{[\exp(\sigma_X^2) - 1]^{3/2}}
\end{aligned} \qquad (6.159)$$

If normal variates X_y^s and X_y^u are used to generate lognormally distributed streamflows Q_y^s 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

$$\rho_Q(k; s, u) = \frac{\exp[\rho_X(k; s, u)\sigma_X(s)\sigma_X(u)] - 1}{\{\exp[\sigma_X^2(s)] - 1\}^{1/2}\{\exp[\sigma_X^2(u)] - 1\}^{1/2}} \qquad (6.160)$$

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 $x_y^s = \ell n(q_y^s - \tau)$ and using the historical correlations of these x_y^s values as estimators of $\rho_X(k; s, u)$ (Stedinger 1981).

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

$$X_{y+1} - \mu = \rho_X(X_y - \mu) + V_y \sigma_X \sqrt{1 - \rho_X^2} \qquad (6.161)$$

then the corresponding Q_y follow the model (Matalas 1967)

$$Q_{y+1} = \tau + D_y\{\exp[\mu_x(1 - \rho_X)]\}(Q_y - \tau)^{\rho_X} \qquad (6.162)$$

where

$$D_y = \exp[(1 - \rho_X^2)^{1/2}\sigma_X V_y] \qquad (6.163)$$

The conditional mean and standard deviation of Q_{y+1} given that $Q_y = q_y$ now depend on $(q_y - \tau)^{\rho_X}$. Because the conditional mean of Q_{y+1} is no longer a linear function of q_y, 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 natural include such differences by employing a Markov chain description of the occurrence of wet and dry days (Wilks 1998).

6.8.5 Multivariate Models

If long concurrent streamflow records can be constructed at the several sites at which synthetic streamflows are desired, then ideally a general multisite streamflow model could be employed. O'Connell (1977), Ledolter (1978), Salas et al. (1980) and Salas (1993) discuss multivariate models and parameter estimation. Unfortunately, model identification (parameter value estimation) is very difficult for the general multivariate models.

This section illustrates how the basic univariate annual flow model in Sect. 8.3 can be generalized to the multivariate case. This exercise reveals how easily multivariate models can be constructed to reproduce specified variances and covariances of the flow vectors of interest, or some transformation of those values. This multisite 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 multisite/multiseason modeling procedures, as is done in the next section employing what have been called disaggregation models. However, while the size of the model matrices and vectors increases, the models are 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 $\mathbf{Z}_y = (Z_y^1, \ldots, Z_y^n)^{\mathrm{T}}$ be the column vector of transformed zero-mean annual flows at sites $s = 1$, 2, …, n, so that

$$E[Z_y^s] = 0 \qquad (6.164)$$

In addition, let $\mathbf{V}_y = \left(V_y^1, \ldots, V_y^n\right)^{\mathrm{T}}$ 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 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

$$\mathbf{Z}_{y+1} = \mathbf{A}\mathbf{Z}_y + \mathbf{B}\mathbf{V}_y \qquad (6.165)$$

where $\mathbf{A}$ and $\mathbf{B}$ 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 $\mathbf{S}_0$ and $\mathbf{S}_1$; the lag-zero covariance matrix, denoted $\mathbf{S}_0$, is defined as

$$\mathbf{S}_0 = E[\mathbf{Z}_y \mathbf{Z}_y^{\mathrm{T}}] \qquad (6.166)$$

and has elements

$$S_0(i,j) = E\left[Z_y^i Z_y^j\right] \qquad (6.167)$$

The lag-one covariance matrix, denoted $\mathbf{S}_1$, is defined as

$$\mathbf{S}_1 = E[\mathbf{Z}_{y+1} \mathbf{Z}_y^{\mathrm{T}}] \qquad (6.168)$$

and has elements

$$S_1(i,j) = E\left[Z_{y+1}^i Z_y^j\right] \qquad (6.169)$$

The covariances do not depend on y because the streamflows are assumed to be stationary.

Matrix $\mathbf{S}_1$ contains the lag 1 covariances and lag 1 cross-covariances. $\mathbf{S}_0$ is symmetric because

the cross covariance $S_0(i, j)$ equals $S_0(j, i)$. In general, $\mathbf{S}_1$ is not symmetric.

The variance–covariance equations that define the values of $\mathbf{A}$ and $\mathbf{B}$ in terms of $\mathbf{S}_0$ and $\mathbf{S}_1$ are obtained by manipulations of Eq. 6.165. Multiplying both sides of that equation by $\mathbf{Z}_y^{\mathrm{T}}$ and taking expectations yields

$$E\left[\mathbf{Z}_{y+1}\mathbf{Z}_y^{\mathrm{T}}\right] = E\left[\mathbf{A}\mathbf{Z}_y\mathbf{Z}_y^{\mathrm{T}}\right] + E\left[\mathbf{B}\mathbf{V}_y\mathbf{Z}_y^{\mathrm{T}}\right] \tag{6.170}$$

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 Eq. 6.170, $E\left[\mathbf{A}\mathbf{Z}_y\mathbf{Z}_y^{\mathrm{T}}\right]$, is a matrix whose (i, j)th element equals

$$E\left[\sum_{k=1}^{n} a_{ik}Z_y^k Z_y^i\right] = \sum_{k=1}^{n} a_{ik}E[Z_y^k Z_y^i] \tag{6.171}$$

The matrix with these elements is the same as the matrix $\mathbf{A}E\left[\mathbf{Z}_y\mathbf{Z}_y^{\mathrm{T}}\right]$.

Hence, $\mathbf{A}$—the matrix of constants—can be pulled through the expectation operator just as is done in the scalar case where $E[aZ_y + b] = aE[Z_y] + b$ for fixed constants a and b.

Substituting $\mathbf{S}_0$ and $\mathbf{S}_1$ for the appropriate expectations in Eq. 6.170 yields

$$\mathbf{S}_1 = \mathbf{A}\mathbf{S}_0 \quad \text{or} \quad \mathbf{A} = \mathbf{S}_1\mathbf{S}_0^{-1} \tag{6.172}$$

A relationship to determine the matrix $\mathbf{B}$ is obtained by multiplying both sides of Eq. 6.165 by its own transpose (this is equivalent to squaring both sides of the scalar equation $a = b$) and taking expectations to obtain

$$E[\mathbf{Z}_{y+1}\mathbf{Z}_{y+1}^{\mathrm{T}}] = E[\mathbf{A}\mathbf{Z}_y\mathbf{Z}_y^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}] + E[\mathbf{A}\mathbf{Z}_y\mathbf{V}_y^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}]$$
$$+ E\left[\mathbf{B}\mathbf{V}_y\mathbf{Z}_y\mathbf{A}^{\mathrm{T}}\right] + E\left[\mathbf{B}\mathbf{V}_y\mathbf{V}_y^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\right] \tag{6.173}$$

The second and third terms on the right-hand side of Eq. 6.173 vanish because the components of Z_y and V_y are independent and have zero mean. $E\left[\mathbf{V}_y\mathbf{V}_y^{\mathrm{T}}\right]$ equals the identity matrix because the components of V_y are independently distributed with unit variance. Thus

$$\mathbf{S}_0 = \mathbf{A}\mathbf{S}_0\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{B}^{\mathrm{T}} \tag{6.174}$$

Solving of the $\mathbf{B}$ matrix one finds that it should satisfy

$$\mathbf{B}\mathbf{B}^{\mathrm{T}} = \mathbf{S}_0 - \mathbf{A}\mathbf{S}_0\mathbf{A}^{\mathrm{T}} = \mathbf{S}_0 - \mathbf{S}_1\mathbf{S}_0^{-1}\mathbf{S}_1^{\mathrm{T}} \tag{6.175}$$

The last equation results from substitution of the relationship for $\mathbf{A}$ given in Eq. 6.172 and the fact that $\mathbf{S}_0$ is symmetric; hence $\mathbf{S}_0^{-1}$ is symmetric.

It should not be too surprising that the elements of $\mathbf{B}$ are not uniquely determined by Eq. 6.175. The components of the random vector $\mathbf{V}_y$ may be combined in many ways to produce the desired covariances as long as $\mathbf{B}$ satisfies Eq. 6.175. A lower triangular matrix that satisfies Eq. 6.175 can be calculated by Cholesky decomposition (Young 1968; Press et al. 1986).

Matalas and Wallis (1976) call Eq. 6.165 the *lag*-1 *model*. They did not call the lag-1 model a Markov model because the streamflows at individual sites do not have the covariances of an autoregressive Markov process given in Eq. 6.153. They suggest an alternative model 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 the covariances of an autoregressive Markov process at each site. In their Markov model, the new $\mathbf{A}$ matrix is simply a diagonal matrix whose diagonal elements are the lag-1 correlations of flows at each site

$$\mathbf{A} = \mathrm{diag}[\rho(1; i, i)] \tag{6.176}$$

where $\rho(1; i, i)$ is the lag-one correlation of flows at site i.

The corresponding $\mathbf{B}$ matrix depends on the new $\mathbf{A}$ matrix and $\mathbf{S}_0$, where as before

$$\mathbf{B}\mathbf{B}^{\mathrm{T}} = \mathbf{S}_0 - \mathbf{A}\mathbf{S}_0\mathbf{A}^{\mathrm{T}} \qquad (6.177)$$

The idea of fitting time series models to each site separately and then correlating in innovations in those separate models to reproduce the cross-correlation between the series is a very general and powerful modeling 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).

6.8.6 Multiseason, Multisite 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.

6.8.6.1 Disaggregation Model

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 level 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 multiseason single-site or multisite streamflow generation. They represent a very flexible modeling 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 $\mathbf{Z}_y = \left(Z_y^1, \ldots, Z_y^N\right)^{\mathrm{T}}$ be the column vector of N transformed normally distributed annual or aggregate annual flows for N separate sites or basins. Next let $\mathbf{X}_y = \left(X_{1,y}^1, \ldots, X_{T,y}^1, X_{1,y}^2, \ldots, X_{Ty}^2, \ldots, X_{1y}^n, \ldots, X_{Ty}^n\right)^{T}$ be the column vector of nT transformed normally distributed seasonal flows X_{ty}^s for season t, year y, and site s.

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

$$\mathbf{X}_y = \mathbf{A}\mathbf{Z}_y + \mathbf{B}\mathbf{V}_y \qquad (6.178)$$

where $\mathbf{V}_y$ is a vector of nT independent standard normal random variables, and $\mathbf{A}$ and $\mathbf{B}$ are, respectively, $nT \times N$ and $nT \times nT$ matrices. One selects values of the elements of $\mathbf{A}$ and $\mathbf{B}$ 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 $\mathbf{A}$ and $\mathbf{B}$ are determined using the matrices $\mathbf{S}_{zz} = E[\mathbf{Z}_y\mathbf{Z}_y^{\mathrm{T}}]$, $\mathbf{S}_{zz} = E[\mathbf{Z}_y\mathbf{Z}_y^{\mathrm{T}}]$,

$\mathbf{S}_{xx} = E[\mathbf{X}_y \mathbf{X}_y^{\mathrm{T}}]$, $\quad \mathbf{S}_{xz} = E[\mathbf{X}_y \mathbf{Z}_y^{\mathrm{T}}]$, and $\quad \mathbf{S}_{zy} = E[\mathbf{Z}_y \mathbf{X}_y^{\mathrm{T}}]$ where $\mathbf{S}_{zz}$ was called $\mathbf{S}_0$ earlier. Clearly, $\mathbf{S}_{xz}^{\mathrm{T}} = \mathbf{S}_{zx}$. If $\mathbf{S}_{xz}$ is to be reproduced, then by multiplying Eq. 6.178 on the right by $\mathbf{Z}_y^{\mathrm{T}}$ and taking expectations, one sees that $\mathbf{A}$ must satisfy

$$E\left[\mathbf{X}_y \mathbf{Z}_y^{\mathrm{T}}\right] = E\left[\mathbf{A}\mathbf{Z}_y \mathbf{Z}_y^{\mathrm{T}}\right] \qquad (6.179)$$

or

$$\mathbf{S}_{xz} = \mathbf{A}\mathbf{S}_{zz} \qquad (6.180)$$

Solving for the coefficient matrix $\mathbf{A}$ one obtains

$$\mathbf{A} = \mathbf{S}_{xz} \mathbf{S}_{zz}^{-1} \qquad (6.181)$$

To obtain an equation that determines the required value of the matrix $\mathbf{B}$, one can multiply both sides of Eq. 6.178 by their transpose and take expectations to obtain

$$\mathbf{S}_{xx} = \mathbf{A}\mathbf{S}_{zz}\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{B}^{\mathrm{T}} \qquad (6.182)$$

Thus to reproduce the covariance matrix $\mathbf{S}_{xx}$ the $\mathbf{B}$ matrix must satisfy

$$\mathbf{B}\mathbf{B}^{\mathrm{T}} = \mathbf{S}_{xx} - \mathbf{A}\mathbf{S}_{zz}\mathbf{A}^{\mathrm{T}} \qquad (6.183)$$

Equations 6.181 and 6.183 for determining $\mathbf{A}$ and $\mathbf{B}$ are completely analogous to Eqs. 6.172 and 6.175 for the $\mathbf{A}$ and $\mathbf{B}$ matrices of the lag-1 models developed earlier. However, for the disaggregation model as formulated, $\mathbf{B}\mathbf{B}^{\mathrm{T}}$ and hence the matrix $\mathbf{B}$ 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 $(T-1)$ 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 nonlinear transformations of the actual flows Q_{ty}^s, then $\mathbf{B}\mathbf{B}^{\mathrm{T}}$ is generally sufficiently non-singular that a $\mathbf{B}$ 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 $\mathbf{A}$ matrix has mn elements. The lower diagonal $\mathbf{B}$ matrix and the symmetric $\mathbf{S}_{xx}$ matrix, upon which it depends, each have $m(m+1)/2$ nonzero or nonredundant elements. For example, when disaggregating two aggregate annual flow series to monthly flows at five sites, $n = 2$ and $m = 12 \times 5 = 60$ so that $\mathbf{A}$ has 120 elements while $\mathbf{B}$ and $\mathbf{S}_{xx}$ each have 1830 nonzero or nonredundant parameters. As the number of sites included in the disaggregation increases, the size of $\mathbf{S}_{xx}$ and $\mathbf{B}$ increases rapidly. Very quickly the model can become over parameterized, and there will be insufficient data to estimate all parameters (Grygier and Stedinger 1988).

In particular, one can think of Eq. 6.178 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 40 years, there would be only 1 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 and not attempting to explicitly reproduce every season-to-season correlation by constructing what have been called condensed and contemporaneous models (Lane 1979; Stedinger and Vogel 1984; Gryier and Stedinger 1988; Koutsoyiannis and Manetas 1996). Condensed models do not attempt to reproduce the cross-correlations among all the flow variates at the same site within a year (Lane 1979; Stedinger et al. 1985), whereas contemporaneous models are like the Markov model developed earlier in Sect. 8.5 and are essentially 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 Eq. 6.177. Grygier and Stedinger (1991) describe how this can be done for a condensed disaggregation model without generating inconsistencies.

6.8.6.2 Aggregation Models

One can start with annual or seasonal flows, and break them down into flows in shorter periods representing months or weeks. Or instead one can start with a model that describes flows and the shortest time step included in the model; this latter approach has been referred to as aggregation model to distinguish it from the disaggregation approach.

One method for generating multiseason 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 modeled 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, it implicitly assumes 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 1962) of monthly flows may be written

$$Z_{t+1,y} = \beta_t Z_{ty} + \sqrt{1 - \beta_t^2}\, V_{ty} \qquad (6.184)$$

where the Z_{ty}'s are standard normal random variables corresponding to the streamflow in season t of year y, β_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 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 persistence than the Thomas-Fiering model. 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, Eq. 6.142 would be

called an AR(1) or AR(1, 0) model. A simple ARMA(1, 1) model is

$$Z_{t+1} = \phi_1 \cdot Z_t + V_{t+1} - \theta_1 \cdot V_t \qquad (6.185)$$

The correlations of this model have the values

$$\rho_1 = (1 - \theta_1\phi_1)(\phi_1 - \theta_1)/(1 + \theta_1^2 - 2\phi_1\theta_1) \qquad (6.186)$$

for the first lag. For $i > 1$

$$\rho_i = \phi^{i-1}\rho_1 \qquad (6.187)$$

For ϕ values near one and $0 < \theta_1 < \phi_1$ the autocorrelations ρ_k can decay much slower than those of the standard AR(1) model.

The correlation function ρ_k of general ARMA (p, q) model

$$Z_{t+1} = \sum_{i=1}^{p} \phi_i \cdot Z_{t+1-i} + V_{t+1} - \sum_{j=1}^{q} \theta_j \cdot V_{t+1-j} \qquad (6.188)$$

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, p. 2553) 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 (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 one (see for example, Bartolini and Salas 1993).

6.9   Stochastic Simulation

This section introduces stochastic simulation. Much more detail on simulation is contained in later parts of this chapter and in the next chapter. Simulation is the most 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 policy, many such alternatives must be simulated.

As with optimization models, simulation models may be deterministic or stochastic. One of the most useful tools in water resource systems planning is stochastic simulation. While optimization can be used to help define reasonable design and operating policy alternatives to be simulated, it takes those simulations to obtain the best insights of just how each such alternative will perform. Stochastic simulation of complex systems on digital computers provides planners with a way to define the probability distribution of performance indices of complex stochastic water resources systems.

When simulating any system, the modeler 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 modeler 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 simulation model and how one can deal with the information that is generated.

6.9.1 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, probability distributions for the variables 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 Sect. 6.3.2).

Most computers have algorithms for generating random numbers uniformly distributed between zero and one. This uniform distribution is defined by its cdf and pdf;

$$F_u(u) = 0 \quad \text{for } u \leq 0,$$
$$u \text{ for } 0 \leq u \leq 1 \quad \text{and} \quad 1 \quad \text{if } u \geq 1 \tag{6.189}$$

and

$$f_U(u) = 1 \quad \text{if } 0 \leq u \leq 1 \quad \text{and} \quad 0 \text{ otherwise} \tag{6.190}$$
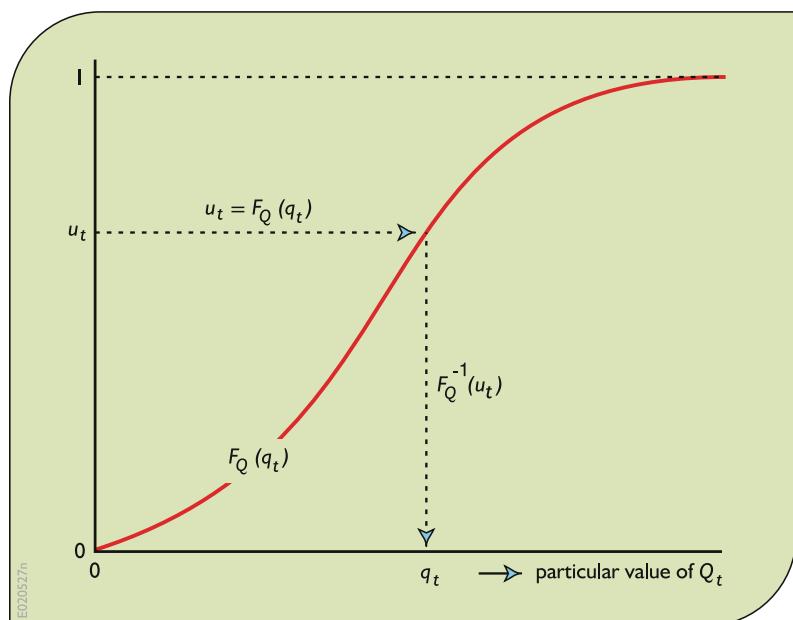
These uniform random variables can then be transformed into random variables with any desired distribution. If $F_Q(q_t)$ is the cumulative distribution function of a random variable Q_t in period t, then Q_t can be generated using the inverse function, as

$$Q_t = F_Q^{-1}[U_t] \tag{6.191}$$

Here U_t is the uniform random number used to generate Q_t. This is illustrated in Fig. 6.13.

Analytical expressions for the inverse of many distributions, such as the normal distribution, are not known, so that special algorithms are

Fig. 6.13 The probability distribution of a random variable can be inverted to produce values of the random variable

employed to efficiently generate deviates with these distributions (Fishman 2001).

6.9.2 River Basin Simulation

An example will demonstrate the use of stochastic simulation in the design and analysis of water resource 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 from 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 4×10^7 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 3×10^7 m^3/yr after construction of the dam to 4×10^7 m^3/yr within 6 years. After that, demand will grow more slowly, to 4.5×10^7 m^3/yr, the estimated maximum reliable yield. The projected demand for summer irrigation water is shown in Table 6.12.

Table 6.12 Projected water demand for irrigation water

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

Table 6.13 Characteristics of the river flow

	winter	summer	annual	
mean flow	4.0	2.5	6.5	$\times 10^7 \, \mathrm{m}^3$
standard deviation	1.5	1.0	2.3	$\times 10^7 \, \mathrm{m}^3$
correlation of flows:				
winter with following summer		0.65		
summer with following winter		0.60		

A simulation study will evaluate how the system can be expected to perform over a 20-year planning period. Table 6.13 contains statistics that describe the hydrology at the dam site. The estimated moments are computed from the 45-year historic record. .

Using the techniques discussed in the previous section, a Thomas-Fiering model is used to generate 25 lognormally distributed synthetic streamflow sequences. The statistical characteristics of the synthetic flows are those listed in Table 6.14. Use of only the 45-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 20-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.

6.9.3   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

$$R_{1y} = \begin{cases} S_{1y} + Q_{1y} - K & \text{if } S_{1y} + Q_{1y} - R_{\min} > K \\ R_{\min} & \text{if } K \geq S_{1y} + Q_{1y} - R_{\min} \geq 0 \\ S_{1y} + Q_{1y} & \text{otherwise} \end{cases}$$
$$(6.192)$$
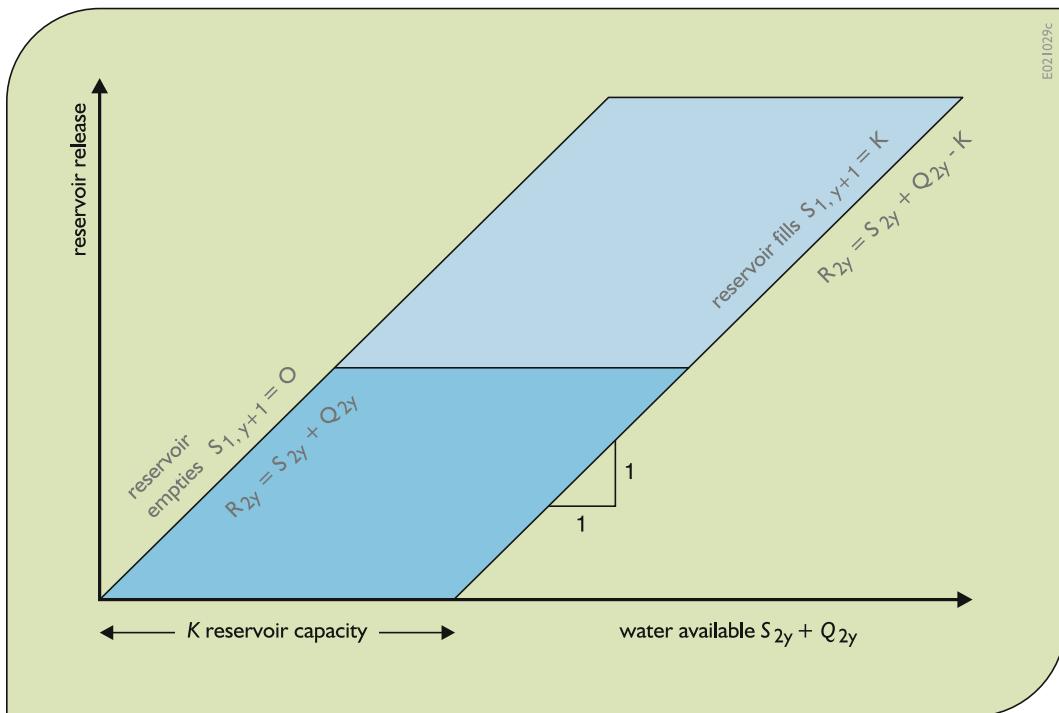
where K is the reservoir capacity of $4 \times 10^7 \, \mathrm{m}^3$ and $R_{\min}$ is $0.50 \times 10^7 \, \mathrm{m}^3$, the minimum release to be made if possible. The volume of water in storage at the beginning of the year's summer season is

$$S_{2y} = S_{1y} + Q_{1y} - R_{1y} \qquad (6.193)$$

The summer release policy is to meet each year's projected demand or target release D_y, if possible, so that

$$\begin{aligned} R_{2y} &= S_{2y} + Q_{2y} - K & \text{if } S_{2y} + Q_{2y} - D_y > K \\ &= D_y & \text{if } 0 \leq S_{2y} + Q_{2y} - D_y \leq K \\ &= S_{2y} + Q_{2y} & \text{otherwise} \end{aligned}$$
$$(6.194)$$

This operating policy is illustrated in Fig. 6.14.

Fig. 6.14 Summer reservoir operating policy. The *shaded area* denotes the feasible region of reservoir releases

The volume of water in storage at the beginning of the next winter season is

$$S_{1,y+1} = S_{2y} + Q_{2y} - R_{2y} \qquad (6.195)$$
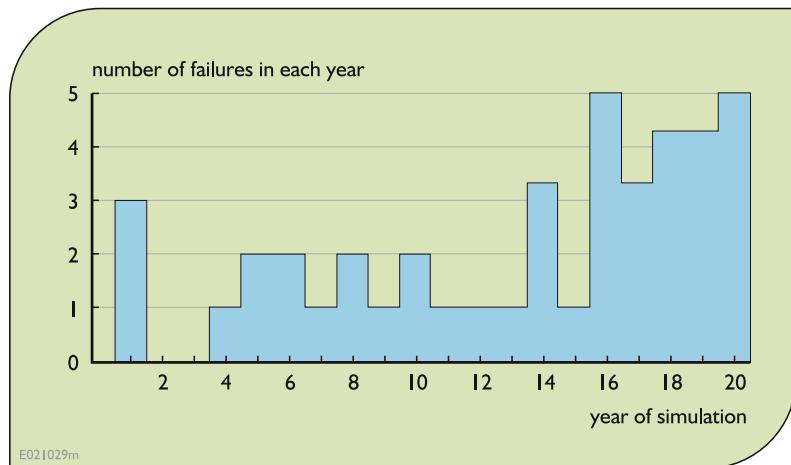
6.9.4 Simulation of the Basin

The question to be addressed by this simulation study is how well will the reservoir 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 resource system. In our 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 next step is to simulate the proposed system to evaluate the specified indices. For our reservoir-irrigation system, the reservoir's operation was simulated 25 times using the 25 synthetic streamflow sequences, each 20 years in length. Each of the 20 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 6.12.

After simulating the system, one must proceed 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.

Fig. 6.15 Number of failures in each year of 25 twenty-year simulations

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:

$$P_y = \Pr[R_{2y} \geq D_y] \qquad (6.196)$$

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 6.15 shows the total number of failures that occurred in each year of the 25 simulations. In 3 of the 25 simulations, 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 results 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 high demand placed on the system. Thus one has a sense for how the reliability of the target releases changes over time.

6.9.5 Interpreting Simulation Output

Table 6.14 contains several summary statistics of the 25 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 20, the number of years simulated. At the bottom of column 2 and the other columns are several statistics that summarize the 25 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 interquartile range $[x_{(6)} - x_{(20)}]$, and/or the range $[x_{(1)} - x_{(25)}]$ of the observations, where $x_{(i)}$ is the ith largest observation. Either set of statistics could be used to describe the center 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 6.14 reports that the mean failure rate for the 25 simulations is 0.084, implying that the average reliability over the 20-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 center 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. Certainly, the farmers are interested not only in knowing the mean or median failure frequency but also the

Table 6.14 Results of 25 20-year simulations

simulation number, i	frequency of failure to meet:		total shortage TS $\times 10^7 m^3$	average deficit, AD
	target	80% of target		
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 $\overline{x}$	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

E021101q

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 a confidence interval within which the observations would fall with some prespecified 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, there is about a 90% probability the actual failure rate 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 that 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

$$P = \frac{n+1-i}{n+1} - \frac{i}{n+1} = \frac{n+1-2i}{n+1} \quad (6.197)$$

In our example, the range $[x_{(1)} - x_{(25)}]$ of the 25 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 $(25 + 1 - 2 \times 6)/(25 + 1) = 54\%$.

Table 6.14 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, nonnegative 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 centers 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,

$$t = \frac{\bar{x} - \mu_x}{s_x/\sqrt{n}} \quad (6.198)$$

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 μ_x. A $100(1 - 2\alpha)\%$ confidence interval for μ_x is, approximately,

$$\bar{x} - t_\alpha \frac{s_x}{\sqrt{n}} \leq \mu_x \leq \bar{x} + t_\alpha \frac{s_x}{\sqrt{n}}$$

or

$$0.084 - t_\alpha\left(\frac{0.081}{\sqrt{25}}\right) \leq \mu_x \leq 0.084 + t_\alpha\left(\frac{0.081}{\sqrt{25}}\right) \quad (6.199)$$

If $\alpha = 0.05$, then $t_\alpha = 1.65$ and Eq. 6.199 becomes $0.057 \leq \mu_x \leq 0.11$.

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 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 6.14 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 9 out of 25, or in 36% of the simulation runs

Columns 4 and 5 of Table 6.14 contain two other indices that pertain to the severity of the failures. The total shortfall in Column 4 is calculated as

$$\text{TS} - \sum_{y-1}^{20} \left[D_{2y} - R_{2y} \right]^+$$

where

$$[Q]^+ = \begin{cases} Q & \text{if } Q > 0 \\ 0 & \text{otherwise} \end{cases} \qquad (6.200)$$

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 is

$$\text{AD} = \frac{1}{m} \sum_{y=1}^{20} \frac{\left[D_{2y} - R_{2y} \right]}{D_{2y}} \qquad (6.201)$$

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 25 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 18 years 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 6.14, 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 rerun 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 6.15 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 25 twenty-year simulations, as opposed to 10 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 figure is misleading because in 14 of the 25

Table 6.15 Results of 25 20-Year simulations with modified operating policy to avoid severe shortages

simulation number, i	frequency of failure to meet:		total shortage TS $\times 10^7 m^3$	average deficit, AD
	target	80% of target		
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

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 19 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 increases while vulnerability decreases. If the farmers are willing to put up with more frequent minor shortages, it appears 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 ith 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 25 paired simulation runs, $y_i < 0$ in 21 cases and $y_i = 0$ in four cases. We can ignore the times 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 21 cases when $y_i \neq 0$ is 2^{-21} or 5×10^{-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 25 simulation runs. However, in all 4 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

$$\bar{y} = \frac{1}{25} \sum_{i-1}^{25} (x_{1i} - x_{2i}) = \bar{x}_1 - \bar{x}_2$$
$$= 0.084 - 0.142 = -0.058 \qquad (6.202)$$

and their variance is

$$s_y^2 = \frac{1}{25} \sum_{i=1}^{25} (x_{1i} - x_{2i} - \bar{y})^2 = (0.0400)^2$$
$$(6.203)$$

Now if the sample size n, equal to 25 here, is sufficiently large, then t defined by

$$t = \frac{\bar{y} - \mu_Y}{s_Y/\sqrt{n}} \qquad (6.204)$$

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 μ_Y equals zero and upon substituting the observed values of $\bar{y}$ and s_Y^2 into Eq. 6.204, one obtains

$$t = \frac{-0.0580}{0.0400/\sqrt{25}} = -7.25 \qquad (6.205)$$

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 μ_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

$$\begin{aligned}
\text{Var}(Y) &= E[X_1 - X_2 - (\mu_1 - \mu_2)]^2 \\
&= E[(X_1 - \mu_1)2] - 2E[(X_1 - \mu_1)(X_2 - \mu_2)] \\
&\quad E[(X_2 - \mu_2)^2] \\
&= \sigma_{x_1}^2 - 2\text{Cov}(X_1, X_2) + \sigma_{X_2}^2
\end{aligned}$$

$$(6.206)$$

where $\text{Cov}(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)]$ 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 $\text{Cov}(X_1, X_2)$ were zero is

$$\hat{\sigma}_Y^2 = s_{x_1}^2 + s_{x_2}^2 = (0.081)^2 + (0.087)^2 = (0.119)^2$$

$$(6.207)$$

The actual sample estimate s_Y equals 0.040; if independent streamflow sequences are used in all simulations, s_Y will take a value near 0.119 rather than 0.040 (Eq. 6.203). A standard deviation of 0.119 yields a value of the test statistic

$$t = \frac{\bar{y} - \mu_Y}{0.119/\sqrt{25}}\Big|_{\mu Y=0} = -2.44 \qquad (6.208)$$

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. 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.

6.10 Conclusions

This chapter has introduced some approaches analysts can consider and use when working with the randomness or uncertainty of their data. Most of the data water resource systems analysts use is uncertain. This uncertainty comes from not understanding as well as we would like how our water resource systems (including its 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 both the supplies of water we manage and 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 some 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 have been widely 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 the 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 chapter is merely an introduction to some of the 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 provided in the next section provide additional and more detailed information.

Although many of the methods presented in this and the following two 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 that 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.

References

Ayyub, B. M., & McCuen, R. H. (2002). *Probability, statistics, and reliability for engineers and scientists*. Boca Raton: Chapman & Hill, CRC Press.

Bartolini, P., & Salas, J. (1993). Modeling of streamflow processes at different time scales. *Water Resources Research, 29*(8), 2573–2587.

Bates, B. C., & Campbell, E. P. (2001). A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research, 37*(4), 937–947.

Beard, L. R. (1960). Probability estimates based on small normal-distribution samples. *Journal of Geophysical Research, 65*(7), 2143–2148.

Beard, L. R. (1997). Estimating flood frequency and average annual damages. *Journal of Water Resources Planning and Management, 123*(2), 84–88.

Benjamin, J. R., & Cornell, C. A. (1970). *Probability, statistics and decisions for civil engineers*. New York: McGraw-Hill Book Company.

Bickel, P. J., & Doksum, K. A. (1977). *Mathematical statistics: Basic ideas and selected topics*. San Francisco: Holden-Day Inc.

Bobée, B. (1975). The log Pearson Type 3 distribution and its applications in hydrology. *Water Resources Research, 14*(2), 365–369.

Bobée, B., & Ashkar, F. (1991). *The gamma distribution and derived distributions applied in hydrology*. Littleton, CO: Water Resources Press.

Bobée, B., & Robitaille, R. (1977). The use of the Pearson Type 3 distribution and Log Pearson Type 3 distribution revisited. *Water Resources Research, 13*(2), 427–443.

Box, G. E. P., Jenkins, G. M., & Risinsel, G. C. (1994). *Times series analysis: Forecasting and control* (3rd ed.). New Jersey: Prentice-Hall Inc.

Camacho, F., McLeod, A. I., & Hipel, K. W. (1985). Contemporaneous autoregressive-moving average (CARMA) modeling in water resources. *Water Resources Bulletin, 21*(4), 709–720.

Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical bayes methods for data analysis* (2nd ed.). New York, NY: Chapman and Hall, CRC.

Chowdhury, J. U., & Stedinger, J. R. (1991). Confidence intervals for design floods with estimated skew coefficient. *Journal of Hydraulic Engineering, 117*(7), 811–831.

Chowdhury, J. U., Stedinger, J. R., & Lu, L.-H. (1991). Goodness-of-fit tests for regional GEV flood distributions. *Water Resources Research, 27*(7), 1765–1776.

Claps, P., Rossi, F., & Vitale, C. (1993). Conceptual-stochastic modeling of seasonal runoff using autoregressive moving average models and different scales of aggregation. *Water Resources Research, 29*(8), 2545–2559.

Cohn, C. A., Lane, W. L., & Stedinger, J. R. (2001). Confidence intervals for EMA flood quantile estimates. *Water Resources Research, 37*(6), 1695–1706.

Cohn, T. A., Lane, W. L., & Baier, W. G. (1997). An algorithm for computing moments-based flood quantile estimates when historical flood information is available. *Water Resources Research, 33*(9), 2089–2096.

Crainiceanu, C. M., Ruppert, D., Stedinger, J. R., & 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.

D'Agostino, R. B., & Stephens, M. A. (1986). *Goodness-of-fit procedures* (p. 560). New York: Marcel Dekker Inc.

David, H. A. (1981). *Order statistics* (2nd ed.). New York: Wiley.

Fiering, M. B. (1967). *Streamflow synthesis*. Cambridge, Mass: Harvard University Press.

Fill, H., & Stedinger, J. (1995). L-moment and PPCC goodness-of-fit tests for the Gumbel distribution and effect of autocorrelation. *Water Resources Research, 31*(1), 225–229.

Fill, H., & Stedinger, J. (1998). Using regional regression within index flood procedures and an empirical Bayesian estimator. *Journal of Hydrology, 210*(1–4), 128–145.

Filliben, J. J. (1975). The probability plot correlation test for normality. *Technometrics, 17*(1), 111–117.

Fishman, G. S. (2001). *Discrete-event simulation: Modeling, programming, and analysis*. Berlin: Springer.

Gabriele, S., & Arnell, N. (1991). A hierarchical approach to regional flood frequency analysis. *Water Resources Research, 27*(6), 1281–1289.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Boca Raton: Chapman & Hall, CRC.

Giesbrecht, F., & Kempthorne, O. (1976). Maximum likelihood estimation in the three-parameter log normal distribution. *Journal of the Royal Statistical Society B, 38*(3), 257–264.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. London, New York: Chapman & Hall.

Greenwood, J. A., & Durand, D. (1960). Aids for fitting the gamma distribution by maximum likelihood. *Technometrics, 2*(1), 55–65.

Griffis, V. W., & Stedinger, J. R. (2007a). The LP3 distribution and its application in flood frequency analysis, 1. Distribution characteristics. *Journal of Hydraulic Engineering, 12*(5), 482–491.

Griffis, V. W., & Stedinger, J. R. (2007b). The LP3 distribution and its application in flood frequency analysis, 2. Parameter estimation methods. *Journal of Hydraulic Engineering, 12*(5), 492–500.

Grygier, J. C., & Stedinger, J. R. (1988). Condensed disaggregation procedures and conservation corrections. *Water Resources Research, 24*(10), 1574–1584.

Grygier, J. C., & Stedinger, J. R. (1991). "SPIGOT: A synthetic flow generation software package, user's manual and technical description," Version 2.6. Ithaca, NY: School of Civil and Environmental Engineering, Cornell University.

Gumbel, E. J. (1958). *Statistics of extremes* (p. 375). New York: Columbia University Press.

Gupta, V. K., & Dawdy, D. R. (1995a). Physical interpretation of regional variations in the scaling exponents of flood quantiles. *Hydrological Processes, 9*(3/4), 347–361.

Gupta, V. K., & Dawdy, D. R. (1995b). Multiscaling and skew separation in regional floods. *Water Resources Research, 31*(11), 2761–2776.

Gupta, V. K., Mesa, O. J., & Dawdy, D. R. (1994). Multiscaling theory of flood peaks: Regional quantile analysis. *Water Resources Research, 30*(12), 3405–3412.

Haan, C. T. (1977). *Statistical methods in hydrology*. Ames, Iowa: Iowa State University Press.

Haas, C. N., & Scheff, P. A. (1990). Estimation of averages in truncated samples. *Environmental Science and Technology, 24*(6), 912–919.

Helsel, D. R. (1990). Less than obvious: Statistical treatment of data below the detection limit. *Environmental Science and Technology, 24*(12), 1767–1774.

Helsel, D. R., & Cohn, T. A. (1988). Estimation of descriptive statistics for multiple censored water quality data. *Water Resources Research, 24*(12), 1997–2004.

Hirsch, R. M. (1979). Synthetic hydrology and water supply reliability. *Water Resources Research, 15*(6), 1603–1615.

Hirsch, R. M., & Stedinger, J. R. (1987). Plotting positions for historical floods and their precision. *Water Resources Research, 23*(4), 715–727.

Hoshi, K., Burges, S. J., & Yamaoka, I. (1978). Reservoir design capacities for various seasonal operational hydrology models. In *Proceedings of the Japanese Society of Civil Engineers* (No. 273, pp. 121–134).

Hoshi, K., Stedinger, J. R., & Burges, S. (1984). Estimation of log normal quantiles: Monte Carlo results and first-order approximations. *Journal of Hydrology, 71*(1/2), 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, 52*(2), 105–124.

Hosking, J. R. M., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics, 29*(3), 339–349.

Hosking, J. R. M., & Wallis, J. R. (1995). A comparison of unbiased and plotting-position estimators of L-moments. *Water Resources Research, 31*(8), 2019–2025.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: An approach based on L-moments*. Cambridge: Cambridge University Press.

Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability weighted moments. *Technometrics, 27*(3), 251–261.

IACWD (Interagency Advisory Committee on Water Data). (1982). *Guidelines for determining flood flow frequency*. Reston Virginia: Bulletin 17B, U.S. Department of the Interior, U.S. Geological Survey, Office of Water Data Coordination.

Jenkins, G. M., & Watts, D. G. (1968). *Spectral analysis and its applications*. San Francisco: Holden-Day, Inc.

Kendall, M. G., & Stuart, A. (1966). *The advanced theory of statistics* (Vol. 3). New York: Hafner Publishing Co., Inc.

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, 8*(5), 1251–1254.

Kirby, W. (1974). Algebraic boundness of sample statistics. *Water Resources Research, 10*(2), 220–222.

Kite, G. W. (1988). *Frequency and risk analysis in hydrology*. Littleton, CO: Water Resources Publications.

Kottegoda, M., & Rosso, R. (1997). *Statistics, probability, and reliability for civil and environmental engineers*. New York: McGraw-Hill.

Koutsoyiannis, D., & Manetas, A. (1996). Simple disaggregation by accurate adjusting procedures. *Water Resources Research, 32*(7), 2105–2117.

Kroll, K., & Stedinger, J. R. (1996). Estimation of moments and quantiles with censored data. *Water Resources Research, 32*(4), 1005–1012.

Kuczera, G. (1982). Combining site-specific and regional information, an empirical Bayesian approach. *Water Resources Research, 18*(2), 306–314.

Kuczera, G. (1983). Effects of sampling uncertainty and spatial correlation on an empirical bayes procedure for combining site and regional information. *The Journal of Hydrology, 65*(4), 373–398.

Kuczera, G. (1999). Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. *Water Resources Research, 35*(5), 1551–1557.

Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1978). Some comparisons of flood statistics in real and log space. *Water Resources Research, 14*(5), 902–920.

Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research, 15*(5), 1055–1064.

Lane, W. (1979). *Applied stochastic techniques (users manual), Bureau of reclamation*. Denver, Colorado: Engineering and Research Center, December.

Langbein, W. B. (1949). Annual floods and the partial duration flood series, EOS. *Transactions of AGU, 30*(6), 879–881.

Ledolter, J. (1978). The analysis of multivariate time series applied to problems in hydrology. *Journal of Hydrology, 36*(3–4), 327–352.

Lettenmaier, D. P., & Burges, S. J. (1977a). Operational assessment of hydrologic models of long-term persistence. *Water Resources Research, 13*(1), 113–124.

Lettenmaier, D. P., & Burges, S. J. (1977b). An operational approach to preserving skew in hydrologic models of long-term persistence. *Water Resources Research, 13*(2), 281–290.

Lettenmaier, D. P., Wallis, J. R., & Wood, E. F. (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research, 23*(2), 313–324.

MacBertouex, P., & Brown, L. C. (2002). *Statistics for environmental engineers* (2nd ed.). Boca Raton, Florida: Lewis Publishers, CRC Press.

Madsen, H., Pearson, C. P., Rasmussen, P. F., & Rosbjerg, D. (1997a). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1. At-site modeling. *Water Resources Research, 33*(4), 747–758.

Madsen, H., Pearson, C. P., & Rosbjerg, D. (1997b). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 2. Regional modeling. *Water Resources Research, 33*(4), 759–770.

Madsen, H., & Rosbjerg, D. (1997a). The partial duration series method in regional index flood modeling. *Water Resources Research, 33*(4), 737–746.

Madsen, H., & Rosbjerg, D. (1997b). Generalized least squares and empirical bayesian estimation in regional partial duration series index-flood modeling. *Water Resources Research, 33*(4), 771–782.

Maheepala, S., & Perera, B. J. C. (1996). Monthly hydrologic data generation by disaggregation. *Journal of Hydrology, 178*, 277–291.

Marco, J. B., Harboe, R., & Salas, J. D. (Eds.). (1989). *Stochastic hydrology and its use in water resources systems simulation and optimization*. Dordrecht: NATO ASI Series, Kluweer Academic Publishers.

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum likelihood GEV quantile estimators for hydrologic data. *Water Resources Research, 36*(3), 737–744.

Martins, E. S., & Stedinger, J. R. (2001a). Generalized maximum likelihood Pareto-Poisson flood risk analysis for partial duration series. *Water Resources Research, 37*(10), 2559–2567.

Martins, E. S., & Stedinger, J. R. (2001b). Historical information in a GMLE-GEV framework with partial duration and annual maximum series. *Water Resources Research, 37*(10), 2551–2557.

Matalas, N. C. (1967). Mathematical assessment of synthetic hydrology. *Water Resources Research, 3*(4), 937–945.

Matalas, N. C., & Wallis, J. R. (1973). Eureka! It fits a Pearson Type 3 distribution. *Water Resources Research, 9*(3), 281–289.

Matalas, N. C., & Wallis, J. R. (1976). Generation of synthetic flow sequences. In A. K. Biswas (Ed.), *Systems approach to water management*. New York: McGraw-Hill Book Company.

Mejia, J. M., & Rousselle, J. (1976). Disaggregation models in hydrology revisited. *Water Resources Research, 12*(2), 185–186.

NERC (Natural Environmental Research Council). (1975). *Flood studies report. Hydrological studies* (Vol. I, 550 pp). London, England.

North, M. (1980). Time-dependent stochastic model of floods. *Journal of the Hydraulics Division, ASCE, 106* (HY5), 649–665.

O'Connell, D. R. H., Ostenaa, D. A., Levish, D. R., & Klinger, R. E. (2002). Bayesian flood frequency analysis with paleohydrologic bound data. *Water Resources Research, 38*(5), 16–1 to 16–4.

O'Connell, P. E. (1977). ARMA models in synthetic hydrology. In T. A. Ciriani, V. Maione, & J. R. Wallis (Eds.), *Mathematical models for surface water hydrology*. New York: Wiley.

Pishro-Nik, H. (2014). Introduction to probability, statistics, and random processes. *Kappa Research.* LLC ISBN/ASIN: 0990637204, ISBN-13: 9780990637202, 744 pp. www.probabilitycourse.com

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986). *Numerical recipes, the art of scientific computing.* Cambridge: Cambridge University Press.

Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory.* Cambridge, Mass: MIT Press.

Rasmussen, P. F., & Rosbjerg, D. (1989). Risk estimation in partial duration series. *Water Resources Research, 25*(11), 2319–2330.

Rasmussen, P. F., & Rosbjerg, D. (1991a). Evaluation of risk concepts in partial duration series. *Stochastic Hydrology and Hydraulics, 5*(1), 1–16.

Rasmussen, P. F., & Rosbjerg, D. (1991b). Prediction uncertainty in seasonal partial duration series. *Water Resources Research, 27*(11), 2875–2883.

Rasmussen, R. F., Salas, J. D., Fagherazzi, L., Rassam, J. C., & Bobee, R. (1996). Estimation and validation of contemporaneous PARMA models for streamflow simulation. *Water Resources Research, 32*(10), 3151–3160.

Robinson, J. S., & Sivapalan, M. (1997). Temporal scales and hydrologic regimes: Implications for flood frequency scaling. *Water Resources Research, 33*(12), 2981–2999.

Rosbjerg, D. (1985). Estimation in partial duration series with independent and dependent peak values. *Journal of Hydrology, 76*, 183–195.

Rosbjerg, D., & Madsen, H. (1998). In H. Wheater & C. Kirby (Eds.), *Design with uncertain design values, hydrology in a changing environment* (Vol. III, pp. 155–163). New York: Wiley.

Rosbjerg, D., Madsen, H., & Rasmussen, P. F. (1992). Prediction in partial duration series with generalized Pareto-distributed exceedances. *Water Resources Research, 28*(11), 3001–3010.

Salas, J. D. (1993). Analysis and modeling of hydrologic time series (Chapter 17). In D. Maidment (Ed.), *Handbook of hydrology.* New York: McGraw-Hill Book Company.

Salas, J. D., Delleur, J. W., Yejevich, V., & Lane, W. L. (1980). *Applied modeling of hydrologic time series.* Littleton, Colorado: Water Resources Press Publications.

Salas, J. D., & Fernandez, B. (1993). Models for data generation in hydrology: Univariate techniques. In J. B. Marco, R. Harboe, & J. D. Salas (Eds.), *Stochastic hydrology and its use in water resources systems simulation and optimization* (pp. 76–95). Dordrecht: Kluwer Academic Publishers.

Salas, J. D., & Obeysekera, J. T. B. (1992). Conceptual basis of seasonal streamflow time series. *Journal of Hydraulic Engineering, 118*(8), 1186–1194.

Schaake, J. C., & Vicens, G. J. (1980). Design length of water resource simulation experiments. *Journal of Water Resources Planning and Management, 106*(1), 333–350.

Slack, J. R., Wallis, J. R., & Matalas, N. C. (1975). On the value of information in flood frequency analysis. *Water Resources Research, 11*(5), 629–648.

Smith, R. L. (1984). Threshold methods for sample extremes. In J. Tiago de Oliveira (Ed.), *Statistical extremes and applications* (pp. 621–638). Dordrecht, Holland: D. Reidel Publishing Company.

Stedinger, J. R. (1980). Fitting log normal distributions to hydrologic data. *Water Resources Research, 16*(3), 481–490.

Stedinger, J. R. (1981). Estimating correlations in multivariate streamflow models. *Water Resources Research, 17*(1), 200–208.

Stedinger, J. R. (1983). Estimating a regional flood frequency distribution. *Water Resources Research, 19* (2), 503–510.

Stedinger, J. R. (1997). Expected probability and annual damage estimators. *Journal of Water Resources Planning and Management, 123*(2), 125–35 [With discussion, Leo R. Beard, *Journal of Water Resources Planning and Management, 124*(6), 365–366, 1998].

Stedinger, J. R. (2000). Flood frequency analysis and statistical estimation of flood risk (Chapter 12). In E. E. Wohl (Ed.), *Inland flood hazards: Human, Riparian and aquatic communities.* Stanford, United Kingdom: Cambridge University Press.

Stedinger, J. R., & Baker, V. R. (1987). Surface water hydrology: Historical and paleoflood information. *Reviews of Geophysics, 25*(2), 119–124.

Stedinger, J. R., & Cohn, T. A. (1986). Flood frequency analysis with historical and paleoflood information. *Water Resources Research, 22*(5), 785–793.

Stedinger, J. R., & Lu, L. (1995). Appraisal of regional and index flood quantile estimators. *Stochastic Hydrology and Hydraulics, 9*(1), 49–75.

Stedinger, J. R., Pei, D., & Cohn, T. A. (1985). A disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations. *Water Resources Research, 21*(5), 665–667.

Stedinger, J. R., & Taylor, M. R. (1982a). Synthetic streamflow generation, part I. Model verification and validation. *Water Resources Research, 18*(4), 919–924.

Stedinger, J. R., & Taylor, M. R. (1982b). Synthetic streamflow generation, part II. Effect of parameter uncertainty. *Water Resources Research, 18*(4), 919–924.

Stedinger, J. R., & Vogel, R. (1984). Disaggregation procedures for the generation of serially correlated flow vectors. *Water Resources Research, 20*(1), 47–56.

Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events (Chapter 18). In D. Maidment (Ed.), *Handbook of hydrology.* New York: McGraw-Hill, Inc.

Tao, P. C., & Delleur, J. W. (1976). Multistation, multiyear synthesis of hydrologic time series by disaggregation. *Water Resources Research, 12*(6), 1303–1311.

Tarboton, D. G., Sharma, A., & Lall, U. (1998). Disaggregation procedures for stochastic hydrology

based on nonparametric density estimation. *Water Resources Research, 34*(1), 107–119.

Tasker, G. D., & Stedinger, J. R. (1986). Estimating generalized skew with weighted least squares regression. *Journal of Water Resources Planning and Management, 112*(2), 225–237.

Thom, H. C. S. (1958). A note on the gamma distribution. *Monthly Weather Review, 86*(4), 117–122.

Thomas, H. A., Jr., & 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, & G. M. Fair (Eds.), *Design of water resources systems*. Cambridge, MA: Harvard University Press.

Thyer, M., & Kuczera, G. (2000). Modeling long-term persistence in hydroclimatic time series using a hidden state markov model. *Water Resources Research, 36* (11), 3301–3310.

Valdes, J. R., Rodriguez, I., & Vicens, G. (1977). Bayesian generation of synthetic streamflows, 2. The multivariate case. *Water Resources Research, 13* (2), 291–295.

Valencia, R., & Schaake, J. C., Jr. (1973). Disaggregation processes in stochastic hydrology. *Water Resources Research, 9*(3), 580–585.

Vicens, G. J., Rodríguez-Iturbe, I., & Schaake, J. C., Jr. (1975). A bayesian framework for the use of regional information in hydrology. *Water Resources Research, 11*(3), 405–414.

Vogel, R. M. (1987). The probability plot correlation coefficient test for the normal, lognormal, and gumbel distributional hypotheses. *Water Resources Research, 22*(4), 587–590 (1986; correction, 23(10), 2013).

Vogel, R. M., & Fennessey, N. M. (1993). L-Moment diagrams should replace product moment diagrams. *Water Resources Research, 29*(6), 1745–1752.

Vogel, R. M., & Shallcross, A. L. (1996). The moving blocks bootstrap versus parametric time series models. *Water Resources Research, 32*(6), 1875–1882.

Vogel, R. M., & Stedinger, J. R. (1988). The value of stochastic streamflow models in over-year reservoir design applications. *Water Resources Research, 24*(9), 1483–1490.

Vogel, R. W., & McMartin, D. E. (1991). Proability plot goodness-of-fit and skewness estimation procedures for the Pearson type 3 distribution. *Water Resources Research*, 27(42), 3149–3158. December.

Wallis, J. R. (1980). Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures. In E. Guggino, G. Rossi, & E. Todini (Eds.), *Piene e Siccita* (pp. 3–36). Catania, Italy: Fondazione Politecnica del Mediterraneo.

Wallis, J. R., Matalas, N. C., & Slack, J. R. (1974a). Just a moment! *Water Resources Research, 10*(2), 211–219.

Wallis, J. R., Matalas, N. C., & Slack, J. R. (1974b). *Just a moment!* Appendix, National Technical Information Service, PB-231 816.

Wang, Q. (2001). A bayesian joint probability approach for flood record augmentation. *Water Resources Research, 37*(6), 1707–1712.

Wang, Q. J. (1997). LH moments for statistical analysis of extreme events. *Water Resources Research, 33*(12), 2841–2848.

Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika, 55*(1), 1–17.

Wilks, D. S. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology, 210*, 178–191.

Wilks, D. S. (2002). Realizations of daily weather in forecast seasonal climate. *Journal of Hydrometeorology, 3*, 195–207.

Young, G. K. (1968). Discussion of "mathematical assessment of synthetic hydrology" by N.C. Matalas, and Reply. *Water Resources Research, 4*(3), 681–683.

Zellner, A. (1971). *An introduction to bayesian inference in econometrics*. New York: Wiley.

Zrinji, Z., & Burn, D. H. (1994). Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology, 13*, 1–21.

Additional References (Further Reading)

Claps, P. (1993). Conceptual basis for stochastic models of monthly streamflows. In J. B. Marco, R. Harboe, & J. D. Salas (Eds.), *Stochastic hydrology and its use in water resources systems simulation and optimization* (pp. 229–235). Dordrecht: Kluwer Academic Publishers.

Kumar, D. N., Lall, U., & Peterson, M. R. (2000). Multisite disaggregation of monthly to daily streamflow. *Water Resources Research, 36*(7), 1823–1833.

Pegram, G. G. S. (1989). Why stochastic hydrology? *Developments in Water Engineering* (Die Siviele Ingenieur in Suid-Afrika, October 1989).

Young, G. K., & Pisano, W. C. (1968). Operational hydrology using residuals. *Journal of the Hydraulics Division, ASCE, 94*(HY4), 909–923.
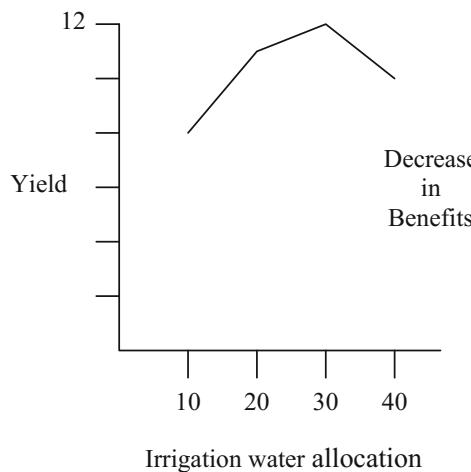
Exercises

6.1 Identify a water resources planning study with which you have some familiarity. Make a list of the basic information used in the study and the methods used transform that information into decisions, recommendations, and conclusions.

(a) Indicate the major sources of uncertainty and possible error in the basic

information and in the transformation of that information into decisions, recommendations, and conclusions.

(b) In systems studies, sources of error and uncertainty are sometimes grouped into three categories

1. Uncertainty due to the natural variability of rainfall, temperature, and stream flows which affect a system's operation.

2. Uncertainty due to errors made in the estimation of the models' parameters with a limited amount of data.

3. Uncertainty or errors introduced into the analysis because conceptual and/or mathematical models do not reflect the true nature of the relationships being described.

Indicate, if applicable, into which category each of the sources of error or uncertainty you have identified falls.

6.2 The following matrix displays the joint probabilities of different weather conditions and of different recreation benefit levels obtained from use of a reservoir in a state park:

Weather	Possible recreation benefits		
	RB_1	RB_2	RB_3
Wet	0.10	0.20	0.10
Dry	0.10	0.30	0.20

(a) Compute the probabilities of recreation levels RB_1, RB_2, RB_3, and of dry and wet weather.

(b) Compute the conditional probabilities P(wet|RB_1), $P(RB_3$|dry), and P(RB_2|wet).

6.3 In flood protection planning, the 100-year flood, which is an estimate of the quantile $x_{0.99}$, is often used as the design flow. Assuming that the floods in different years are independently distributed

(a) Show that the probability of at least one 100-year flood in a 5-year period is 0.049.

(b) What is the probability of at least one 100-year flood in a 100-year period?

(c) If floods at 1000 different sites occur independently, what is the probability of at least one 100-year flood at some site in any single year?

6.4 The price to be charged for water by an irrigation district has yet to be determined. Currently it appears as if there is as 60% probability that the price will be $10 per unit of water and a 40% probability that the price will be $5 per unit. The demand for water is uncertain. The estimated probabilities of different demands given alternative prices are as follows:

Price/Quantity	Prob. of quantity demanded given price				
	30	55	80	100	120
$5	0.00	0.15	0.30	0.35	0.20
$10	0.20	0.30	0.40	0.10	0.00

(a) What is the most likely value of future revenue from water sales?

(b) What are the mean and variance of future water sales?

(c) What is the median value and interquartile range of future water sales?

(d) What price will maximize the revenue from the sale of water?

6.5 Plot the following data on possible recreation losses and irrigated agricultural yields. Show that use of the expected storage level or expected allocation underestimates the expected value of the convex function describing reservoir losses while it overestimates the expected value of the concave function describing
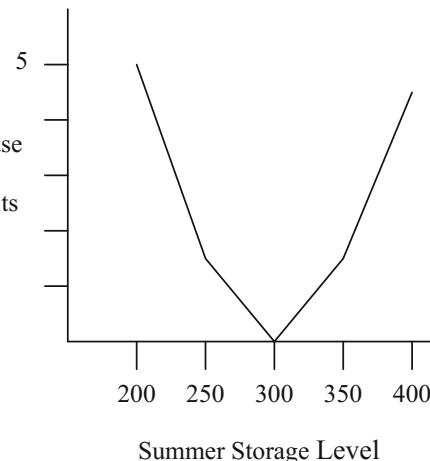
crop yield. A concave function $f(x)$ has the property that $f(x) \le f(x_0) + f'(x_0)(x - x_0)$ for any x_0; prove that use of $f(E[X])$ will always overestimate the expected value of a concave function $f(X)$ when X is a random variable.

Irrigation water allocation	Crop yield/Hectare	Probability of allocation
10	6.5	0.2
20	10	0.3
30	12	0.3
40	11	0.2
Summer storage level	Decrease in recreation benefits	Probability of storage level
200	5	0.1
250	2	0.2
300	0	0.4
350	1	0.2
400	4	0.1

discount rate is high and the life is relatively long, the uncertainty may not very important. However, if the life of a reservoir, or of a wastewater treatment facility, or any other such project, relatively short, the practice of using the expected life to calculate present costs or benefits may be misleading. In this problem, assume that a project results in $1000 of net benefits at the end of each year is expected to last between 10 and 30 years. The probability of ending at the end of each year within the range of 11–30 is the same. Given a discount rate of 10%

(a) Compute the present value of net benefits NB_0, assuming a 20-year project life.
(b) Compare this with the expected present net benefits $E[NB_0]$ taking account of uncertainty in the project lifetime.
(c) Compute the probability that the actual present net benefits is at least $1000 less than NB_0, the benefit esti-

Irrigation water allocation

Summer Storage Level

6.6 Complications can be added to the economic evaluation of a project by uncertainty concerning the usefulness life of the project. For example, the time at which the useful life of a reservoir will end due to silting is never known with certainty when the reservoir is being planned. If the

mate based on a 20-year life.
(d) What is the chance of getting $1000 more than the original estimate NB_0?

6.7 A continuous random variable that could describe the proportion of fish or other animals in different large samples which

have some distinctive features is the beta distribution whose density is ($a > 0, b > 0$)

$$f_X(x) = \begin{cases} cx^{\alpha-1}(1-x)^{\beta-1} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Directly calculate the value of c and the mean and variance of X for $\alpha = \beta = 2$.

(b) In general, $c = \Gamma(\alpha + \beta)/\Gamma(\alpha)\Gamma(\beta)$, where $\Gamma(\alpha)$ is the gamma function equal to $(\alpha - 1)!$ for integer α. Using this information, derive the general expression for the mean and variance of X. To obtain a formula which gives the values of the integrals of interest, note that the expression for c must be such that the integral over (0, 1) of the density function is unity for any α and β.

6.8 The joint probability density of rainfall at two places on rainy days could be described by

$$f_{X,Y}(x,y) \begin{cases} 2/(x+y+1)^3 & x, y \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

Calculate and graph

(a) $F_{XY}(x, y)$, the joint distribution function of X and Y.

(b) $F_Y(y)$, the marginal cumulative distribution function of Y, and $f_Y(y)$, the density function of Y.

(c) $f_{Y|X}(y|x)$, the conditional density function of Y given that $X = x$, and $F_{Y|X}(y|x)$, the conditional cumulative distribution function of Y given that $X = x$ (the cumulative distribution function is obtained by integrating the density function).
Show that

$$F_{Y|X}(y|x=0) > F_Y(y) \quad \text{for } y > 0$$

Find a value of x_0 and y_0 for which

$$F_{Y|X}(y_0|x_0) < F_Y(y_0)$$

6.9 Let X and Y be two continuous independent random variables. Prove that

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

for any two real-valued functions g and h. Then show that $\text{Cov}(X, Y) = 0$ if X and Y are independent.

6.10 A frequent problem is that observations (X, Y) are taken on such quantities as flow and concentration and then a derived quantity $g(X, Y)$ such as mass flux is calculated. Given that one has estimates of the standard deviations of the observations X and Y and their correlation, an estimate of the standard deviation of $g(X, Y)$ is needed. Using a second-order Taylor series expansion for the mean of $g(X, Y)$ as a function of its partial derivatives and of the means, variances, covariance of the X and Y. Using a first-order approximation of g (X, Y), obtained an estimates of the variances of $g(X, Y)$ as a function of its partial derivatives and the moments of X and Y. Note, the covariance of X and Y equals

$$E[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY}^2$$

6.11 A study of the behavior of water waves impinging upon and reflecting off a breakwater located on a sloping beach was conducted in a small tank. The height (crest-to-trough) of the waves was measured a short distance from the wave generator and at several points along the beach different distances from the breakwater were measured and their mean and standard error recorded.

Location	Mean wave height (cm)	Standard error of mean (cm)
Near wave generator	3.32	0.06
1.9 cm from breakwater	4.42	0.09
1.9 cm from breakwater	2.59	0.09
1.9 cm from breakwater	3.26	0.06

At which points were the wave heights significantly different from the height near wave generator assuming that errors were independent?

Of interest to the experimenter is the ratio of the wave heights near the breakwater to the initial wave heights in the deep water. Using the results in Exercise 6.10, estimate the standard error of this ratio at the three points assuming that errors made in measuring the height of waves at the three points and near the wave generator are independent. At which point does the ratio appear to be significantly different from 1.00?

Using the results of Exercise 6.10, show that the ratio of the mean wave heights is probably a biased estimate of the actual ratio. Does this bias appear to be important?

6.12 Derive Kirby's bound, Eq. 6.45, on the estimate of the coefficient of skewness by computing the sample estimates of the skewness of the most skewed sample it would be possible to observe. Derive also the upper bound $(n-1)^{1/2}$ for the estimate of the population coefficient of variation $\frac{v_x}{\bar{x}}$ when all the observations must be nonnegative.

6.13 The errors in the predictions of water quality models are sometimes described by the double exponential distribution whose density is

$$f(x) = \frac{\alpha}{2}\exp(-\alpha|x-\beta|)$$
$$-\infty < x < +\infty$$

What are the maximum likelihood estimates of α and β? Note that

$$\frac{\mathrm{d}}{\mathrm{d}\beta}|x-\beta| = \begin{cases} -1 & x > \beta \\ +1 & x < \beta \end{cases}$$

Is there always a unique solution for β?

6.14 Derive the equations that one would need to solve to obtain maximum likelihood estimates of the two parameters α and β of the gamma distribution. Note an analytical expression for $\mathrm{d}\Gamma(\alpha)/\mathrm{d}\alpha$ is not available so that a closed-form expression for maximum likelihood estimate of α is not available. What is the maximum likelihood estimate of β as a function of the maximum likelihood estimate of α?

6.15 The log-Pearson Type-III distribution is often used to model flood flows. If X has a log-Pearson Type-III distribution then

$$Y = \ln(X) - m$$

has a two-parameter gamma distribution where e^m is the lower bound of X if $\beta > 0$ and e^m is the upper bound of X if $\beta < 0$. The density function of Y can be written

$$f_Y(y)\mathrm{d}y = \frac{(\beta y)^{\alpha-1}}{\Gamma(\alpha)}\exp(-\beta y)\mathrm{d}(\beta y)$$
$$0 < \beta y < +\infty$$

Calculate the mean and variance of X in terms of α, β and m. Note that

$$E[X^r] = E[(\exp(Y+m))^r]$$
$$= \exp(rm)E[\exp(rY)]$$

To evaluate the required integrals remember that the constant terms in the definition of $f_Y(y)$ ensure that the integral of this density function must be unity for any values of α and β so long as $\alpha > 0$ and $\beta y > 0$. For what values of r and β does the mean of X fail to exist? How do the values of m, α and β affect the shape and scale of the distribution of X?

6.16 When plotting observations to compare the empirical and fitted distributions of streamflows, or other variables, it is necessary to assign a cumulative probability to each observation. These are called *plotting positions*. As noted in the text, for the ith largest observation X_i,

$$E[F_X(X_i)] = i/(n+1)$$

Thus the Weibull plotting position $i/(n + 1)$ is one logical choice. Other commonly used plotting positions are the *Hazen plotting position* $(i - 3/8)/(n + 1/4)$. The plotting position $(i - 3/8)/(n + 1/4)$ is a reasonable choice because its use provides a good approximation to the expected value of X_i. In particular for standard normal variables

$$E[X_i] \cong \Phi^{-1}[(i - 3/8)/(n + 1/4)]$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal variable. While much debate centers on the appropriate plotting position to use to estimate $p_i = F_X(X_i)$, often people fail to realize how imprecise all such estimates must be. Noting that

$$\mathrm{Var}(p_i) = \frac{i(n - i - 1)}{(n+1)^2(n+2)},$$

contrast the difference between the estimates $\hat{p}_i$ of p_i provided by these three plotting positions and the standard deviation of p_i. Provide a numerical example. What do you conclude?

6.17 The following data represent a sequence of annual flood flows, the maximum flow rate observed each year, for the Sebou River at the Azib Soltane gaging station in Morocco.

Date	Maximum discharge (m^3/s)	Date	Maximum discharge (m^3/s)
03/26/33	445	03/13/54	750
12/11/33	1410	02/27/55	603
11/17/34	475	04/08/56	880
03/13/36	978	01/03/57	485
12/18/36	461	12/15/58	812
12/15/37	362	12/23/59	1420
04/08/39	530	01/16/60	4090
02/04/40	350	01/26/61	376
02/21/41	1100	03/24/62	904
02/25/42	980	01/07/63	4120
12/20/42	575	12/21/63	1740
02/29/44	694	03/02/65	973
12/21/44	612	02/23/66	378
12/24/45	540	10/11/66	827
05/15/47	381	04/01/68	626
05/11/48	334	02/28/69	3170
05/11/49	670	01/13/70	2790
01/01/50	769	04/04/71	1130
12/30/50	1570	01/18/72	437
01/26/52	512	02/16/73	312
01/20/53	613		

(a) Construct a histogram of the Sebou flood flow data to see what the flow distribution looks like.

(b) Calculate the mean, variance, and sample skew. Based on Table 6.3, does the sample skew appear to be significantly different from zero?

(c) Fit a normal distribution to the data and use the Kolmogorov–Smirnov test to determine if the fit is adequate. Draw a quantile-quantile plot of the fitted quantiles $F^{-1}[(i - 3/8)/(n + 1/4)]$ versus the observed quantiles x_i and include on the graph the

Kolmogorov–Smirnov bounds on each x_i, as shown in Figs. 6.2a, b.

(d) Repeat part (c) using a two-parameter lognormal distribution.

(e) Repeat part (c) using a three-parameter lognormal distribution. The Kolmogorov–Smirnov test is now approximate if applied to $\log_e[X_i - \tau]$, where τ is calculated using Eq. 6.81 or some other method of your choice.

(f) Repeat part (c) for two- and three-parameter versions of the gamma distribution. Again, the Kolmogorov-Smirnov test is approximate.

(g) A powerful test of normality is provided by the correlation test. As described by Filliben (1975), one should approximate $p_i = F_X(x_i)$ by

$$\hat{p}_i = \begin{cases} 1 - (0.5)^{1/n} & i = 1 \\ (i - 0.3175)/(n + 0.365) & i = 2, \ldots, n-1 \\ (0.5)^{1/n} & i = n \end{cases}$$

Then one obtains a test for normality by calculation of the correlation r between the ordered observations X_i and m_i the median value of the ith largest observation in a sample of n standard normal random variables so that

$$m_i = \Phi^{-1}(\hat{p}_i)$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. The value of r is then

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2(m_i - \bar{m})^2}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(m_j - \bar{m})^2}}$$

Some significance levels for the value of r are (Filliben 1975)

	Significance level		
n	1%	5%	10%
10	0.876	0.917	0.934
20	0.925	0.950	0.960
30	0.947	0.964	0.970
40	0.958	0.972	0.977
50	0.965	0.977	0.981
60	0.970	0.980	0.983

The probability of observing a value of r less than the given value, where the observations actually drawn from a normal distribution, equals the specified probability. Use this test to determine whether a normal or two-parameter lognormal distribution provides an adequate model for these flows.

6.18 A small community is considering the immediate expansion of its wastewater treatment facilities so that the expanded facility can meet the current deficit of 0.25 MGD and the anticipated growth in demand over the next 25 years. Future growth is expected to result in the need of an additional 0.75 MGD. The expected demand for capacity as a function of time is

$$\text{Demand} = 0.25\text{MGD} + G(1 - e^{-0.23t})$$

where t is the time in years and $G = 0.75$ MGD. The initial capital costs and maintenance and operating costs related to capital are $\$1.2 \times 10^6 \, C^{0.70}$ where C is the plant capacity (MGD). Calculate the loss Primary>Loss of economic efficiency LEE and the misrepresentation of minimal costs (MMC) that would result if a designer incorrectly assigned G a value of 0.563 or 0.938 ($\pm 25\%$) when determining the required capacity of the treatment

plant. [Note: When evaluating the true cost of a nonoptimal design which provides insufficient capacity to meet demand over a 25-year period, include the cost of building a second treatment plant; use an interest rate of 7% per year to calculate the present value of any subsequent expansions.] In this problem, how important is an error in G compared to an error in the elasticity of costs equal to 0.70? One MGD, a million gallons per day, is equivalent to 0.0438 m^3/s.

6.19 A municipal water utility is planning the expansion of their water acquisition system over the next 50 years. The demand for water is expected to grow and is given by

$$D = 10t(1 - 0.006t)$$

where t is the time in years. It is expected that two pipelines will be installed along an acquired right-of-way to bring water to the city from a distant reservoir. One pipe will be installed immediately and then a second pipe when the demand just equals the capacity C in year t is

$$PV = (\alpha + \beta C^\gamma)e^{-rt}$$

where

$$\alpha = 29.5$$
$$\beta = 5.2$$
$$\gamma = 0.5$$
$$r = 0.07/\text{year}$$

Using a 50-year planning horizon, what is the capacity of the first pipe which minimizes the total present value of the construction of the two pipelines? When is the second pipe built? If a $\pm 25\%$ error is made in estimating γ or r, what are the losses of economic efficiency (LEE) and the misrepresentation of minimal costs (MMC)? When finding the optimal decision with each set of parameters, find the time of the second expansion to the nearest year; a computer program that finds the total present value of costs as a function of the time of the second expansion t for $t = 1, \ldots, 50$ would be helpful. (A second pipe need not be built.)

6.20 A national planning agency for a small country must decide how to develop the water resources of a region. Three development plans have been proposed, which are denoted d_1, d_2, and d_3. Their respective costs are 200f, 100f, and 100f where f is a million farths, the national currency. The national benefits which are derived from the chosen development plan depend, in part, on the international market for the goods and agricultural commodities that would be produced. Consider three possible international market outcomes, m_1, m_2, and m_3. The national benefits if development plan 1 selected would be, respectively, 400, 290, 250. The national benefits from selection of plan 2 would be 350, 160, 120, while the benefits from selection of plan 3 would be 250, 200, 160.

(a) Is any plan inferior or dominated?
(b) If one felt that probabilities could not be assigned to m_1, m_2, and m_3 but wished to avoid poor outcomes, what would be an appropriate decision criterion, and why? Which decisions would be selected using this criterion?
(c) If $\Pr[m_1] = 0.50$ and $\Pr[m_2] = \Pr[m_3] = 0.25$, how would each of the

expected net benefits and expected regret criteria rank the decisions?

6.21 Show that if one has a choice between two water management plans yielding benefits X and Y, where X is stochastically smaller than Y, then for any reasonable utility function, plan Y is preferred to X.

6.22 A reservoir system was simulated for 100 years and the average annual benefits and their variance were found to be

$$\overline{B} = 4.93$$

$$s_B^2 = 3.23$$

The correlation of annual benefits was also calculated and is:

k	r_k
0	1.000
1	0.389
2	0.250
3	0.062
4	0.079
5	0.041

(a) Assume that $\rho(l) = 0$ for $l > k$, compute (using Eq. 6.137) the standard error of the calculated average benefits for $k = 0, 1, 2, 3, 4,$ and 5. Also calculate the standard error of the calculated benefits, assuming that annual benefits may be thought of as a stochastic process with a correlation structure $\rho_B(k) = [\rho_B(1)]^k$. What is the effect of the correlation structure among the observed benefits on the standard error of their average?

(b) At the 90 and 95% levels, which of the r_k are significantly different from zero, assuming that $\rho_B(l) = 0$ for $l > k$?

6.23 Replicated reservoir simulations using two operating policies produced the following results:

Replicate	Benefits	
	Policy 1	Policy 2
1	6.27	4.20
2	3.95	2.58
3	4.49	3.87
4	5.10	5.70
5	5.31	4.02
6	7.15	6.75
7	6.90	4.21
8	6.03	4.13
9	6.35	3.68
10	6.95	7.45
11	7.96	6.86
Mean, X_i	6.042	1.570
Standard deviation of values, s_{xi}	1.217	4.859

(a) Construct a 90% confidence limits for each of the two means $\overline{X}_i$.

(b) With what confidence interval can you state that Policy 1 produces higher benefits than Policy 2 using the sign test and using the t-test?

(c) If the corresponding replicate with each policy were independent, estimate with what confidence one could have concluded that Policy 1 produces higher benefits with the t-test.

6.24 Assume that annual streamflow at a gaging site have been grouped into three categories or states. State 1 is 5–15 m^3/s, state 2 is 15–25 m^3/s, and state 3 is 25–35 m^3/s, and these grouping contain all the flows on records. The following transition probabilities have been computed from record:

P_{ij}		j		
		1	2	3
i	1	0.5	0.3	0.2
	2	0.3	0.3	0.4
	3	0.1	0.5	0.4

(a) If the flow for the current year is between 15 and 25 m^3/s, what is the probability that the annual flow 2 years from now will be in the range 25–35 m^3/s?

(b) What is the probability of a dry, an average, and a wet year many years from now?

6.25 A Markov chain model for the streamflows in two different seasons has the following transition probabilities

Streamflow in Season 1	Streamflow next Season 2		
	0–3 m^3/s	3–6 m^3/s	≥6 m^3/s
0–10 m^3/s	0.25	0.50	0.25
≥10 m^3/s	0.05	0.55	0.40

Streamflow in Season 2	Streamflow next Season 1	
	0–10 m^3/s	≥10 m^3/s
0–3 m^3/s	0.70	0.30
3–6 m^3/s	0.50	0.50

Calculate the steady-state probabilities of the flows in each interval in each season.

6.26 Can you modify the deterministic discrete DP reservoir operating model to include the uncertainty, expressed as P_{ij}^t, of the inflows, as in Exercise 6.25?

(*Hints*: The operating policy would define the release (or final storage) in each season as a function of not only the initial storage but also the inflow. If the inflows change, so might the release or final storage volume. Hence you need to discretize the inflows as well as the storage volumes. Both storage and inflow are state variables. Assume for this model you can predict with certainty the inflow in each period at the beginning of the period. So, each node of the network represents a known initial storage and inflow value. You cannot predict with certainty the following period's flows, only their probabilities. What does the network look like now?

6.27 Assume that there exist two possible discrete flows Q_{it} into a small reservoir in each of two periods t each year having probabilities P_{it}. Find the steady-state operating policy (release as a function of initial reservoir volumes and current period's inflow) for the reservoir that minimizes the expected sum of squared deviations from storage and release targets. Limit the storage volumes to integer values that vary from 3 to 5. Assume a storage volume target of 4 and a release target of 2 in each period t. (Assume only integer values of all states and decision variables and that each period's inflow is known at the beginning of the period.) Find the annual expected sum of squared deviations from the storage and release targets.

Period, t	Flows, Q_{it}		Probabilities, P_{it}	
	$i = 1$	$i = 2$	$i = 1$	$i = 2$
1	1	2	0.17	0.83
2	3	4	0.29	0.71

This is an application of Exercise 6.26 except the flow probabilities are independent of the previous flow.

6.28 Assume that the streamflow Q at a particular site has cumulative distribution function $F_Q(q) = q/(1 + q)$ for $q \geq 0$. Show how to compute the mean streamflow, and the probability that any specified value of streamflow, q, will be exceeded.

6.29 Assume that a potential water user can withdraw water from an unregulated stream, and that the probability distribution function $F_Q()$ of the available streamflow Q is known. Calculate the value of the withdrawal target T that will maximize the expected net benefits from the water's use given the two short-run benefit functions specified below.

(a) The benefits from streamflow Q when the target is T are

$$B(Q|T) = \begin{cases} B_0 + \beta T + \gamma(Q - T) & Q \geq T \\ B_0 + \beta T + \delta(Q - T) & Q < T \end{cases}$$

where $\delta > \beta > \gamma$. In this case, the optimal target T^* can be expressed as a function of $P^* = F_Q(T) = \Pr\{Q \leq T\}$, the probability that the random streamflow Q will be less than or equal to T. Prove that

$$P^* = (\beta - \gamma)/(\delta - \gamma).$$
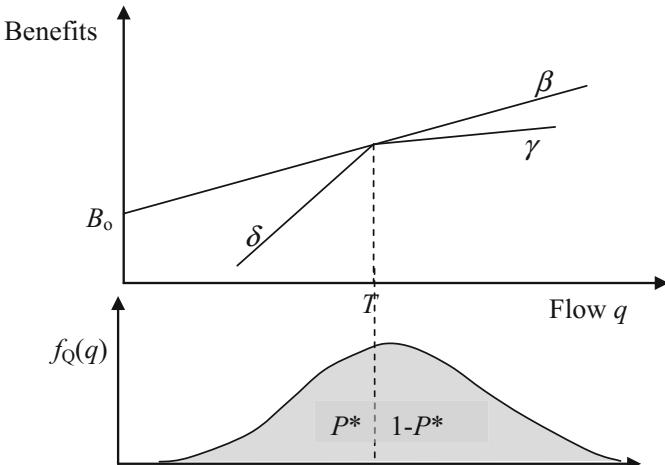
(b) The benefits from streamflow Q when the target is T are

$$B(Q|T) = B_0 + \beta T - \delta(Q - T)^2$$

6.30 If a random variable is discrete, what effect does this have on the specified confidence of a confidence interval for the median or any other quantile? Give an example.

6.31 (a) Use Wilcoxon test for unpaired samples to test the hypothesis that the distribution of the total shortage TS in Table 6.14 is stochastically less than the total shortage TS reported in Table 6.15. Use only the data from the second 10 simulations reported in the table. Use the

fact that observations are paired (i.e., simulation j for $11 \leq j \leq 20$ in both tables were obtained with the same streamflow sequence) to perform the analysis with the sign test.

(b) Use the sign test to demonstrate that the average deficit with Policy 1 (Table 6.14) is stochastically smaller than with Policy 2 (Table 6.15); use all simulations.

6.32 The accompanying table provides an example of the use of non-parametric statistics for examining the adequacy of synthetic streamflow generators. Here the maximum yield that can be supplied with a given size reservoir is considered. The following table gives the rank of the maximum yield obtainable with the historic flows among the set consisting of the historic yield and the maximum yield achievable with 1000 synthetic sequences of 25 different rivers in North America.

(a) Plot the histogram of the ranks for reservoir sizes $S/\mu_Q = 0.85$, 1.35, 2.00. (*Hint*: Use the intervals 0–100, 101–200, 201–300, etc.) Do the ranks look uniformly distributed?

Rank of the Maximum Historic Yield among 1000 Synthetic Yields

River number	Normalized active storage, S/μ_Q			
	0.35	0.85	1.35	2.00
1	47	136	128	235
2	296	207	183	156
3	402	146	120	84
4	367	273	141	191
5	453	442	413	502
6	76	92	56	54
7	413	365	273	279
8	274	191	86	51
9	362	121	50	29
10	240	190	188	141
11	266	66	60	118
12	35	433	562	738
13	47	145	647	379
14	570	452	380	359
15	286	392	424	421
16	43	232	112	97
17	22	102	173	266
18	271	172	260	456
19	295	162	272	291
20	307	444	532	410
21	7	624	418	332
22	618	811	801	679
23	1	78	608	778
24	263	902	878	737
25	82	127	758	910

Source A.I. McLeod and K.W. Hipel, Critical Drought Revisited, Paper presented at the international Symposium on Risk and Reliability in Water Resources, Waterloo, Ont., June 26–28, 1978

   (b) Do you think this streamflow generation model produces streamflows
       which are consistent with the historic
       flows when one uses as a criterion the
       maximum possible yield? Construct a
       statistical test to support your conclusion and show that it does support
       your conclusion. (*Idea*: You might
       want to consider if it is equally likely
       that the rank of the historical yield is

500 and below 501 and above. You
could then use the binomial distribution to determine the significance of
the results.)

   (c) Use the Kolmogrov–Smirnov test to
       check if the distribution of the yields
       obtainable with storage $S/\mu_Q = 1.35$ is
       significantly different from uniform
       $F_U(u) = u$ for $0 \le u \le 1$.   How
       important do you feel this result is?

6.33 Section 7.3 dismisses the bias in v_x^2 for
      correlated X's as unimportant to its variance.

   (a) Calculate the approximate bias in v_x^2
       for   the   cases   corresponding   to
       Table 6.10 and determine if this
       assertion is justified.
   (b) By numerically evaluating the bias
       and variance of v_x^2, when $n = 25$,
       determine if the same result holds if
       $\rho_x(k) = 0.5(0.9)^k$, which is the autocorrelation function of an ARMA(1,
       1) process sometimes used to describe
       annual streamflow series.

6.34 Consider the crop irrigation problem in
      Exercise 4.31. For the given prices 30 and
      25 for crop A and B, the demand for each
      crop varies over time. Records of demands
      show for crop A the demand ranges from 0
      to 10 uniformly. There is an equal probability of that the demand will be any value
      between 0 and 10. For crop B the demand
      ranges from 5 units to 15 units, and the
      most likely demand is 10. At least 5 units
      and no more than 15 units of crop B will
      be demanded. The demand for crop B can
      be defined by a triangular density function,
      beginning with 5, having a mean of 10 and
      an upper limit of 15. Develop and solve a
      model for finding the maximum expected
      net revenue from both crops, assuming the
      costs of additional resources are 2/unit of
      water, 4/unit of land, 8/unit of fertilizer,

and 5/unit of labor. The cost of borrowed money, i.e., the borrowing interest rate, is 8 percent per growing season. How does the solution change if the resource costs are 1/10th of those specified above?

6.35 In Sect. 6.9.2 generated synthetic streamflows sequences were used to simulate a reservoir's operation. In the example, a Thomas-Fiering model was used to generate $\ln(Q_{1y})$ and $\ln(Q_{2y,})$ the logarithms of the flows in the two seasons of each year y, so as to preserve the season-to-season correlation of the untransformed flows. Noting that the annual flow is the sum of the untransformed seasonal flows Q_{1y} and Q_{2y}, calculate the correlation of annual flows produced by this model. The required data are given in Table 6.13. (*Hint*: You need to first calculate the covariance of $\ln(Q_{1y})$ and $\ln(Q_{1,y+1})$ and then of Q_{1y} and $Q_{2,y+1}$).

6.36 Part of New York City's municipal water supply is drawn from three parallel reservoirs in the upper Delaware River basin. The covariance matrix and lag-1 covariance matrix, as defined in Eqs. 6.166 and 6.168, were estimated based on the 50-year flow record to be (in m³/s):

$$S_0 = \begin{bmatrix} 20.002 & 21.436 & 6.618 \\ 21.436 & 25.141 & 6.978 \\ 6.618 & 6.978 & 2.505 \end{bmatrix} = [\mathrm{Cov}(Q_y^i, Q_y^j)]$$

$$S_1 = \begin{bmatrix} 6.487 & 6.818 & 1.638 \\ 7.500 & 7.625 & 1.815 \\ 2.593 & 2.804 & 0.6753 \end{bmatrix} = [\mathrm{Cov}(Q_{y+1}^i, Q_y^j)]$$

Other statistics of the annual flow are

Site	Reservoir	Mean flow	Standard deviation	r_1
1	Pepacton	20.05	4.472	0.3243
2	Cannosville	23.19	5.014	0.3033
3	Neversink	7.12	1.583	0.2696

(a) Using these data, determine the values of the **A** and **B** matrices of the lag 1 model defined by Eq. 6.165. Assume

that the flows are adequately modeled by a normal distribution. A lower triangular **B** matrix that satisfies $M = \mathbf{BB}^T$ may be found by equating the elements of $\mathbf{BB}^T$ to those of M as follows:

$$M_{11} = b_{11}^2 \rightarrow b_{11} = \sqrt{M_{11}}$$

$$M_{21} = b_{11}b_{21} \rightarrow b_{21} = \frac{M_{21}}{b_{11}} = \frac{M_{21}}{\sqrt{M_{11}}}$$

$$M_{31} = b_{11}b_{31} \rightarrow b_{31} = \frac{M_{31}}{b_{11}} = \frac{M_{31}}{\sqrt{M_{11}}}$$

$$M_{22} = b_{21}^2 + b_{22}^2 \rightarrow b_{22}^2 = \sqrt{M_{22} - b_{21}^2}$$

$$= \sqrt{M_{22} - M_{21}^2/M_{11}}$$

and so forth for M_{23} and M_{33}. Note that $b_{ij} = 0$ for $i < j$ and M must be symmetric because $\mathbf{BB}^T$ is necessarily symmetric.

(b) Determine **A** and $\mathbf{BB}^T$ for the Markov model which would preserve the variances and cross-covariances of the flows at each site, but not necessarily the lag 1 cross covariances of the flows. Calculate the lag 1 cross-covariances of flows generated with your calculated **A** matrix.

(c) Assume that some model has been built to generate the total annual flow into the three reservoirs. Construct and calculate the parameters of a disaggregation model that, given the total annual inflow to all three reservoirs, will generate annual inflows into each of the reservoirs preserving the variances and cross-covariances of the flows. [*Hint*: The necessary statistics of the total flows can be calculated from those of the individual flows.]

6.37 Derive the variance of an ARMA(1, 1) process in terms of ϕ_1, θ_1, and σv^2. [*Hint*: Multiply both sides of the equation to obtain a second. Be careful to remember which V_t's are independent of which Z_t's.]

6.38 The accompanying table presents a 60-year flow record for the normalized

flows of the Gota River near Sjotop-Vannersburg in Sweden.

(a) Fit an autoregressive Markov model to the annual flow record.

(b) Using your model, generate a 50-year synthetic flow record. Demonstrate the mean, variance, and correlation of your generated flows deviate from the specified values no more than would be expected as a result of sampling error.

(c) Calculate the autocorrelations and partial autocovariances of the annual flows for a reasonable number of lags. Calculate the standard errors of the calculated values. Determine reasonable value of p and q for an ARMA(p, q) model of the flows. Determine the parameter values for the selected model.

Annual Flows, Gota River near Sjotop-Vannersburg, Sweden

1898	1.158	1918	0.948	1938	0.892
1899	1.267	1919	0.907	1939	1.020
1900	1.013	1920	0.991	1940	0.869
1901	0.935	1921	0.994	1941	0.772
1902	0.662	1922	0.701	1942	0.606
1903	0.950	1923	0.692	1943	0.739
1904	1.120	1924	1.086	1944	0.813
1905	0.880	1925	1.306	1945	1.173
1906	0.802	1926	0.895	1946	0.916
1907	0.856	1927	1.149	1947	0.880
1908	1.080	1928	1.297	1948	0.601
1909	0.959	1929	1.168	1949	0.720
1910	1.345	1930	1.218	1950	0.955
1911	1.153	1931	1.209	1951	1.186
1912	0.929	1932	0.974	1952	1.140
1913	1.158	1933	0.834	1953	0.992
1914	0.957	1934	0.638	1954	1.048
1915	0.705	1935	0.991	1955	1.123
1916	0.905	1936	1.198	1956	0.774
1917	1.000	1937	1.091	1957	0.769

Source V.M. Yevdjevich, Fluctuations of Wet and Dry Years, Part I, Hydrology Paper No. 1, Colorado State University, Fort Collins, Colo., 1963

(d) Using the estimated model in (c), generate a 50-year synthetic streamflow record and demonstrate that the mean, variance, and show that first autocorrelations of the synthetic flows deviate from the modeled values by no more than would be expected as a result of sampling error.

6.39 (a) Assume that one wanted to preserve the covariance matrices $\mathbf{S}_0$ and $\mathbf{S}_1$ of the flows at several site $\mathbf{Z}_y$ using the multivariate or vector ARMA(0, 1) model

$$\mathbf{Z}_{y+1} = \mathbf{A}\mathbf{V}_y - \mathbf{B}\mathbf{V}_{y-1}$$

where $\mathbf{V}_y$ contains n independent standard normal random variables. What is the relationship between the values of $\mathbf{S}_0$ and $\mathbf{S}_1$ and the matrices $\mathbf{A}$ and $\mathbf{B}$?

(b) Derive estimates of the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ of the multivariate AR(2) model

$$\mathbf{Z}_{y+1} = \mathbf{A}\mathbf{Z}_y + \mathbf{B}\mathbf{Z}_{y-1} + \mathbf{C}\mathbf{V}_y$$

using the covariance matrices $\mathbf{S}_0$, $\mathbf{S}_1$, and $\mathbf{S}_2$.

6.40 Create a model for the generation of monthly flows. The generated monthly flows should have the same marginal distributions as were fitted to the observed flows of record and should reproduce (i) the month-to-month correlation of the flows, (ii) the month-to-season correlation between each monthly flow and the total flow the previous season, and (iii) the month-to-year correlation between each monthly flow and the total 12-month flow in the previous year. Show how to estimate the model's parameters. How many parameters does your model have? How are the values of the seasonal model? How do you think this model could be improved?

Decision-makers are increasingly willing to consider the uncertainty associated with model predictions of the economic, environmental, or social impacts associated with 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 of input parameters and variables used in optimization and simulation models can help in quantifying the uncertainty in the resulting model output. This chapter outlines and illustrates some approaches for doing this.

7.1 Introduction

Water resource 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 objectives water serve tend to change over time, and these changes cannot always be predicted. Many of the parameters of models used to predict the multiple hydrologic, economic, environmental, ecological, and social impacts are also uncertain. Indeed, models used to predict these impacts are, at least in part, based on many uncertain assumptions. This uncertainty associated with planning and managing cannot be avoided (WWAP 2012).

To the extent that probabilities can be included where appropriate in models and their inputs at least some of the uncertainty of their outputs can be identified and quantified. 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 impact the decisions that should be made. In other situations it will. Sensitivity analyses can help guide efforts needed to reduce that uncertainty. Model sensitivity and uncertainty analysis are discussed in more detail in the next chapter.
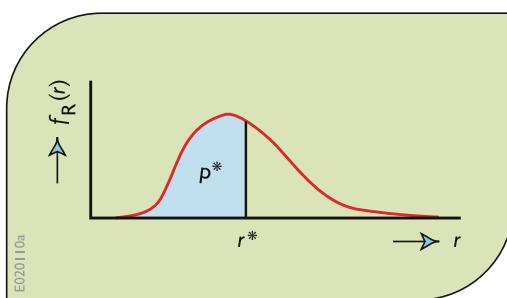
This chapter is divided into two main sections. The first section introduces a number of approaches to probabilistic optimization and simulation modeling. Probabilistic models will be developed and applied to some of the same water resources management problems used to illustrate deterministic modeling in previous chapters. These modeling methods could 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 these commonly used approaches to probabilistic modeling and show how they can be applied to water resources system design and operating problems.

7.2   Generating Values from Known Probability Distributions

As discussed in the previous chapter, variables whose values cannot be predicted with certainty are called random variables. Often inputs to hydrologic simulation models are observed or synthetically generated values of rainfall or streamflow. Other examples of such random variables could be evaporation losses, point and nonpoint source wastewater discharges, demands for water, spot prices for energy that may impact the amount of hydropower to produce, etc. Random processes are considered stationary if the statistical attributes of the process are not changing. If there is no serial correlation in the spatial or temporal sequence of observed values, then such stationary random processes can be characterized by single probability distributions. These probability distributions are often based on past observations of the values of the random variable. These past 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 Fig. 7.1. 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^* in Fig. 7.1. The entire area under a probability density distribution, as shown in Fig. 7.1, is 1.

Fig. 7.1 Probability density distribution of a random variable R. The probability that r is less than or equal r^* is p^*

Integrating this function over the entire range of r, converts the density function to a cumulative distribution function, $F_R(r^*)$, ranging from 0 to 1, as illustrated in Fig. 7.2.

$$\int_0^{r^*} f_R(r)dr = \Pr(r^* \leq R) = F_R(r^*) \qquad (7.1)$$

Given any value of p^* from 0 to 1, one can find its corresponding random variable value r^* from the inverse of the cumulative distribution function.

$$F_R^{-1}(p^*) = r^* \qquad (7.2)$$

From the distribution shown in Fig. 7.1 it is obvious that the likelihood of different values of the random variable varies; values 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 Eq. 7.2, together with a series of uniformly distributed (all equally likely) values of p^* over the range from 0 to 1 (i.e., along the vertical axis of Fig. 7.2), a corresponding series of random variable values, r^*, associated with any distribution can be generated. These random variable values will have a cumulative distribution as shown in Fig. 7.2, and hence a density distribution as shown in Fig. 7.1, regardless of the types or shapes of those distributions. The mean and variance of the distributions will be maintained.
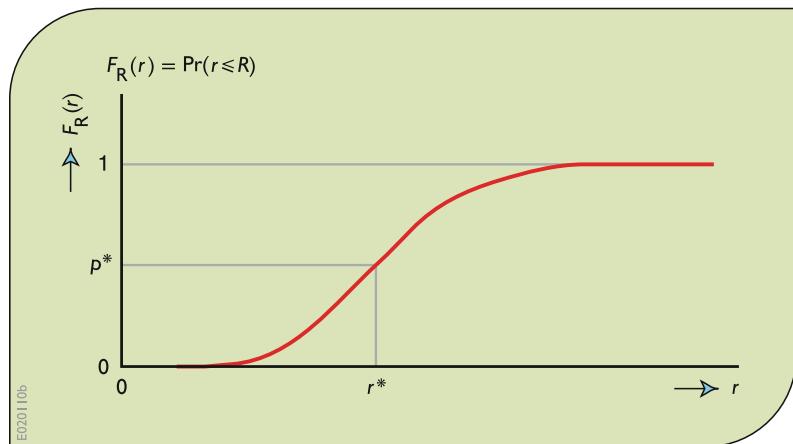
The mean and variance of continuous distributions are

$$\int r f_R(r)dr = E[R] \qquad (7.3)$$

$$\int (r - E[R])^2 f_R(r)dr = \text{Var}[R] \qquad (7.4)$$

The mean, variance and serial correlations of discrete distributions having possible values denoted by r_i with probabilities p_i are

Fig. 7.2 Cumulative distribution function of a random variable R showing the probability of any observed random 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^*

$$\sum_i r_i p_i = E[R] \qquad (7.5)$$

$$\sum_i (r_i - E[R])^2 p_i = \text{Var}[R] \qquad (7.6)$$

If a time series of T random variable values, r_t, generated from the distribution of the same stationary random variable, R, exist, then the serial or autocorrelations of r_t and r_{t+k} in this time series for any positive integer k are

$$\rho_R(k) = \sum_{t=1,T-k} [(r_t - E[R])(r_{t+k} - E[R])] / \sum_{t=1,T} (r_t - E[R])^2 \qquad (7.7)$$

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 ρ. If there is no correlation (ρ is 0), the expected value of R_{t+1} is the mean of the population, $E[R]$. If there is perfect correlation (ρ is 1), 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
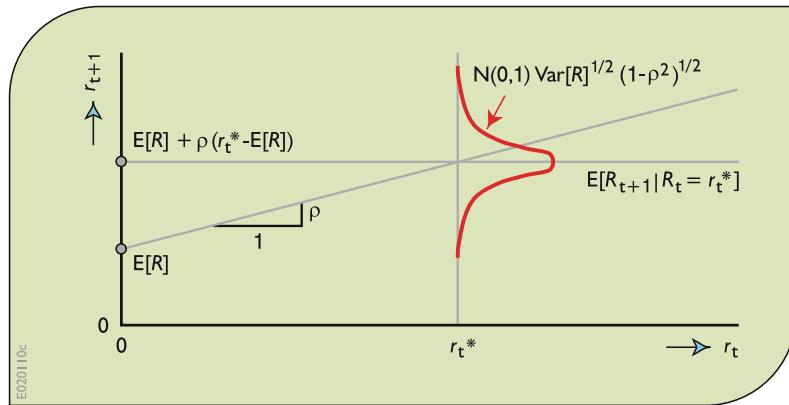
$$E[R_{t+1} | R_t = r_t] = E[R] + \rho(r_t - E[R]). \qquad (7.8)$$

The variance of the random variable R_{t+1} depends on the variance of the distribution, Var $[R]$, and the lag one correlation coefficient, ρ.

$$\text{Var}[R_{t+1} | R_t = r_t] = Var[R](1 - \rho^2). \qquad (7.9)$$

If there is perfect correlation ($\rho = 1$) the process is deterministic and there is no variance. The value for r_{t+1} is r_t. If there is no correlation, i.e., serial correlation does not exist ($\rho = 0$), 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

Fig. 7.3 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

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, $Var[R]$, and lag one correlation ρ, can be obtained from Eq. 7.10.

$$r_{t+1} = E[R] + \rho(r_t - E[R]) + N(0,1)\sigma(1 - \rho^2)^{1/2}.$$
(7.10)

The term $N(0, 1)$ in Eq. 7.10 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 Eq. 7.2 for a $N(0, 1)$ distribution to obtain a value of random number for use in Eq. 7.10. This positive or negative number is substituted for the term $N(0, 1)$ in Eq. 7.10 to obtain a value r_{t+1}. This is shown on the graph in Fig. 7.3.
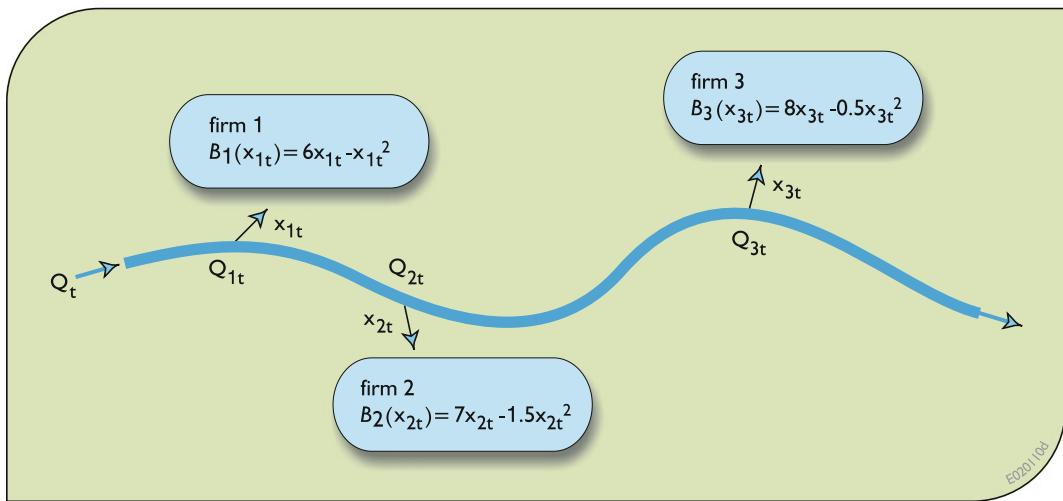
Simulation models that have random inputs, such as a series of r_t values, will generally produce random outputs. After many simulations, the probability distributions of each random output variable value can be defined. These then can be used to estimate reliabilities and other statistical characteristics of those output distributions. This process of generating multiple random inputs for multiple simulations to obtain multiple random outputs is called Monte Carlo simulation.

7.3 Monte Carlo Simulation
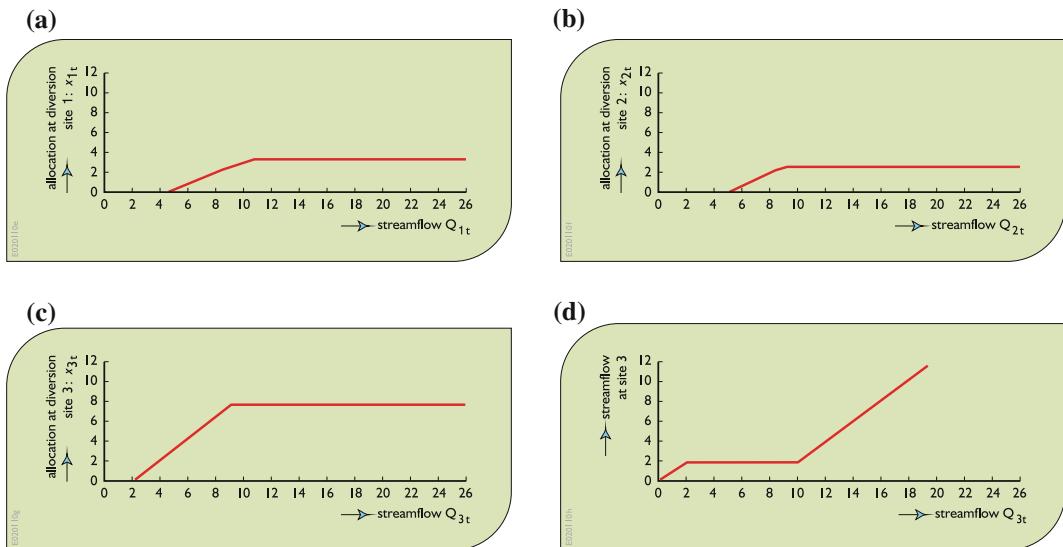
To illustrate Monte Carlo simulation, consider the allocation problem involving three firms, each of which receives a benefit, $B_i(x_{it})$, from the amount of water, x_{it}, allocated to it in each period t. This situation is shown in Fig. 7.4. Monte Carlo simulation can be used to find the probability 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 3 units to Firm 3, and the next 4 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 Fig. 7.5 illustrate this policy. Each allocation plot reflects the priorities given to the three firms and the users further downstream.

A simulation model can now 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 Fig. 7.2 using Eq. 7.2. 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 based on that reduced streamflow, in accordance with the allocation

Fig. 7.4 Stream flow 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
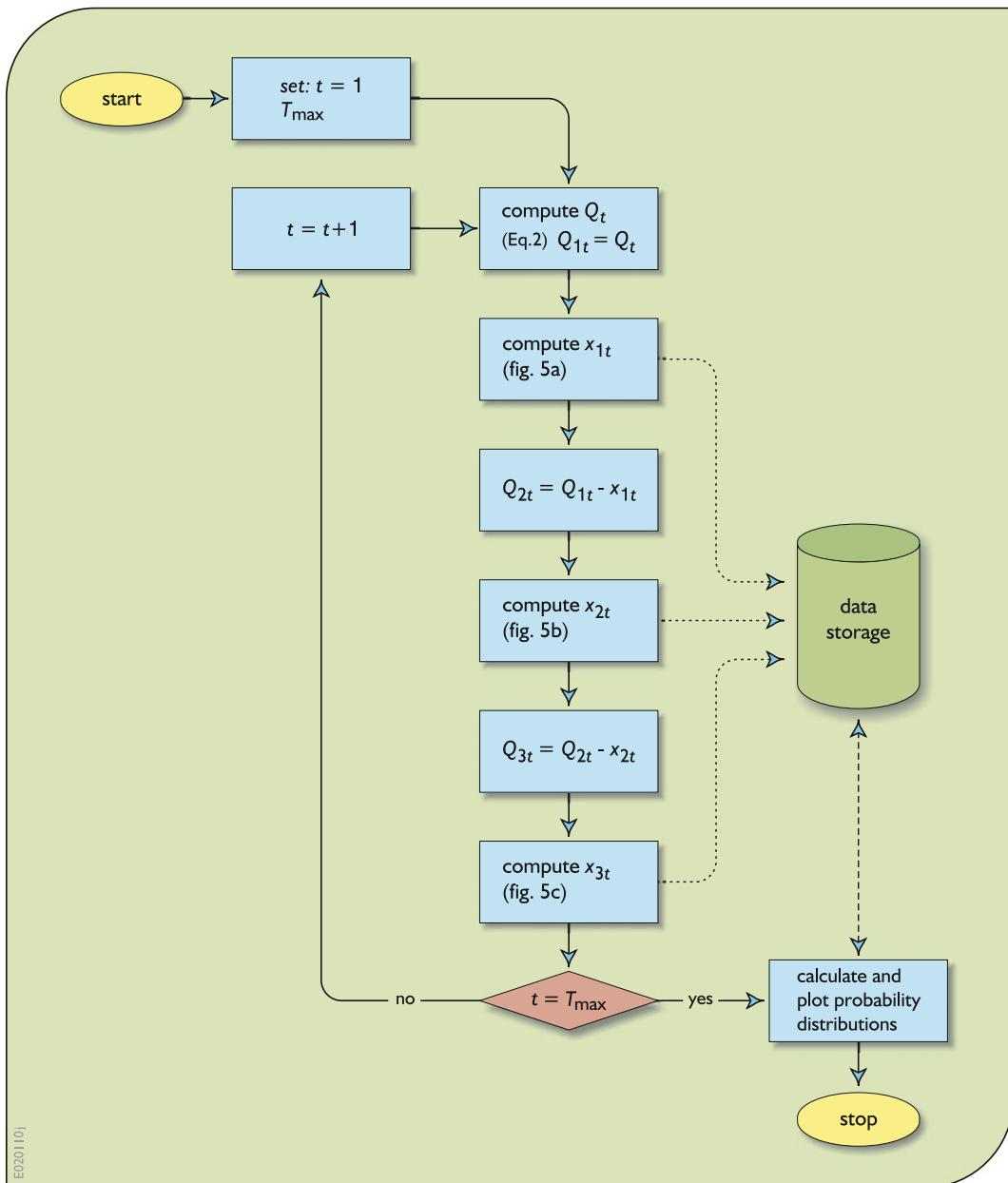
Fig. 7.5 a Water allocation policy for Firm 1 based on the flow at its diversion site. This policy applies for each period t. **b** Water allocation policy for Firm 2 based on the flow at its diversion site for that firm. This policy applies for each period t. **c** Water allocation policy for Firm 3 based on the flow at its diversion site. This policy applies for each period t. **d** Streamflow downstream of site 3 given the streamflow Q_{3t} at site 3 before the diversion. This applies for each period t

policy defined in Fig. 7.5. After numerous time steps the probability distributions of the allocations to each of the firms can be defined.

Figure 7.6 shows a flow chart for this simulation model.

Having defined the probability distribution of the allocations, based on the allocation policy, one can now consider each of the allocations as random variables, X_1, X_2, and X_3 for firms 1, 2 and 3, respectively.

Fig. 7.6 Monte Carlo simulation to determine probability distributions of allocations to each of three water users, as illustrated in Fig. 7.4. The dash lines represent information (data) flows

7.4   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 Fig. 7.4. For planning purposes, the three firms may want to set allocation targets, not expecting

to have those targets met 100% of the time. To insure, 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

$$\Pr\{T_i \leq X_i\} \geq 0.90 \quad i = 1, 2, \text{ and } 3 \quad (7.11)$$
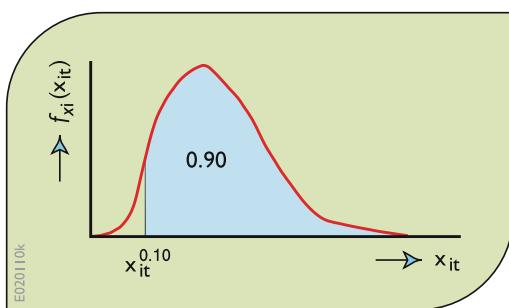
In this constraint, the allocation target T_i is an unknown decision variable, and X_i is a random variable whose distribution has just been computed and is known.

To include chance constraints in optimization models, their deterministic equivalents must be defined. The deterministic equivalents of these three chance constraints in Eq. 7.11 are

$$T_i \leq x_{it}^{0.10} \quad i = 1, 2, \text{ and } 3 \quad (7.12)$$

where $x_{it}^{0.10}$ is the particular value of the random variable X_i that is equaled or exceeded 90% of the time. This value is shown on the probability distribution for X_i in Fig. 7.7.

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 labor used in the firm is presumably based on the target allocations. Once the target is set assume there are no benefits gained by excess allocations of water. 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%.

$$\text{Maximize} \left(6T_1 - T_1^2\right) + \left(7T_2 - 1.5T_2^2\right) + \left(8T_3 - 0.5T_3^2\right) \quad (7.13)$$

Subject to:

$$\Pr\{T_1 + T_2 + T_3 \leq [Q_t - \min(Q_t, 2)]\} \geq 0.90$$
$$\text{for all periods } t \quad (7.14)$$
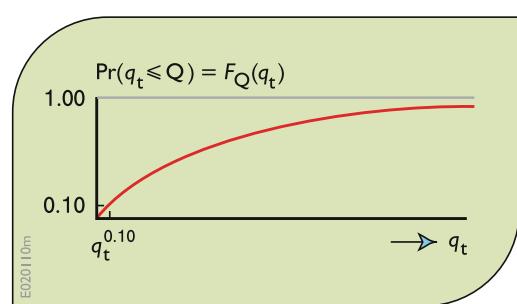
where Q_t is the random streamflow variable upstream of all diversion sites. If the same unconditional probability distribution of Q_t applies for each period t then only one Eq. 7.14 is needed.

Assuming the value of the streamflow, $q_t^{0.10}$, that is equaled or exceeded 90% of the time, is greater than 2 (the amount that must remain in the stream), the deterministic equivalent of chance constraint Eq. 7.14 is

$$T_1 + T_2 + T_3 \leq \left[q_t^{0.10} - \min\left(q_t^{0.10}, 2\right)\right] \quad (7.15)$$

The value of the flow that is equaled or exceeded 90% of the time, $q_t^{0.10}$, can be obtained from the cumulative distribution of flows as illustrated in Fig. 7.8.

Assume this 90% reliable flow is 8. The deterministic equivalent of the chance constraint Eq. 7.14 for all periods t is simply $T_1 + T_2 + T_3 \leq 6$. The optimal solution of the chance

Fig. 7.7 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 equaled or exceeded, as indicated by the shaded region

Fig. 7.8 Example cumulative probability distribution showing the particular value of the random variable, $q_t^{0.10}$, that is equaled or exceeded 90% of the time

constrained target allocation model, Eqs. 7.13 and 7.15, 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.

7.5 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 7.1.

The estimated mean, variance and correlation coefficient of the observed flows shown in Table 7.1 can be calculated using Eqs. 7.16, 7.17 and 7.18.

$$E[Q] = \sum_{t=1}^{31} q_t/31 = 3.155 \qquad (7.16)$$

$$\text{Var}[Q] = \sum_{t=1}^{31} (q_t - 3.155)^2/31 = 1.95 \quad (7.17)$$

Lag-one correlation coefficient $= \rho$
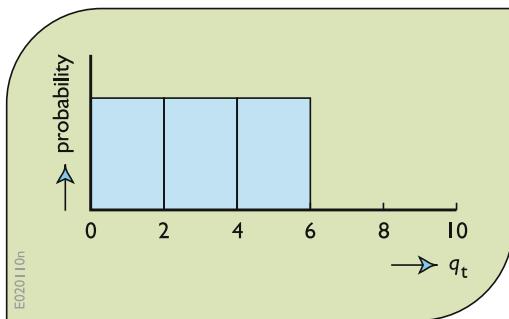
$$= \left[\sum_{t=1}^{30} (q_{t+1} - 3.155)(q_t - 3.155) \right] / \sum_{t=1}^{31} (q_t - 3.155)^2$$

$$= 0.50$$

$$(7.18)$$

The probability distribution of the flows in Table 7.1 can be approximated by a histogram. Histograms can be created by subdividing the entire range of random variable values, e.g., flows, into discrete intervals. For example, let each interval be 2 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 Fig. 7.9. In this case, just to compare this with what will be calculated later, the first flow, q_1, is ignored.

Table 7.1 Sequence of flows for 31 time periods t

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

Fig. 7.9 Histogram showing an equal 1/3 probability that the values of the random variable Q_t will be in any one of the three 2-flow unit intervals

Figure 7.9 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.

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 7.2. The numbers in the matrix are based on the flows in Table 7.1 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 7.2 can be divided by the total number of flow transitions in that row (the sum 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 10 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}.

$$P_{ij} = \Pr\{Q_{t+1} \text{ in interval } j | Q_t \text{ in interval } i\} \tag{7.19}$$

These conditional or transition probabilities, shown in Table 7.3, correspond to the number of transitions shown in Table 7.2.

Table 7.3 is a matrix of transition probabilities. The sum of the probabilities in each row equals 1. Matrices of transition probabilities whose rows sum to one 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 previous flows. Each row would have

Table 7.2 Matrix showing the number of times a flow in interval i in period t was followed by a flow in interval j in period $t + 1$

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

Table 7.3 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

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

the same probabilities as the unconditional distribution shown in Fig. 7.9. 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 7.1 are positively correlated, as indeed has already determined from Eq. 7.18.

Using the information in Table 7.3, 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 $t + 1$ are the probabilities shown in the third row of the matrix in Table 7.3.

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 $t = 3$.

$$\Pr\{Q_3 \text{ in interval } j\} = PQ_{j,3}$$
$$= \sum_i \Pr\{Q_2 \text{ in interval } i\} \Pr\{Q_3 \text{ in interval } j | Q_2 \text{ in interval } i\}$$

(7.20)

The last term in Eq. 7.20 is the transition probability, from Table 7.3, 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 Eqs. 7.19, 7.20 can be written in a general form as

$$PQ_{j,t+1} = \sum_i PQ_{it} P_{ij}$$
for all intervals j and periods t

(7.21)

This operation can be continued to any future time period. Table 7.4 illustrates the results of

Table 7.4 Probabilities of observing a flow in any flow interval i in a future time period t given a current flow in interval $i = 3$

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}

These probabilities are derived using the transition probabilities P_{ij} in Table 7.3 in Eq. 7.21 and assuming the flow interval observed in period 1 is in interval 3

such calculations for 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 Fig. 7.9. 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, Eq. 7.21 can be written as

$$PQ_j = \sum_i PQ_i \, P_{ij} \quad \text{for all intervals } j \text{ less one}$$

$$(7.22)$$

Equation 7.22 along with Eq. 7.23 can be used to calculate all the unconditional probabilities PQ_i directly.

$$\sum_i PQ_i = 1 \qquad (7.23)$$

Conditional or transition probabilities can be incorporated into stochastic optimization models of water resource systems.

7.6 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 discrete 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 $x \le q$ is binding ($x = q$), 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, e.g., 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 known flow q_i is an unknown allocation x_i. Now the deterministic constraint $x \le q$ can be replaced with the set of constraints $x_i \le q_i$ and the term $B(x)$ in the original objective function can be replaced by its expected value, $\sum_i PO_i B(x_i)$.
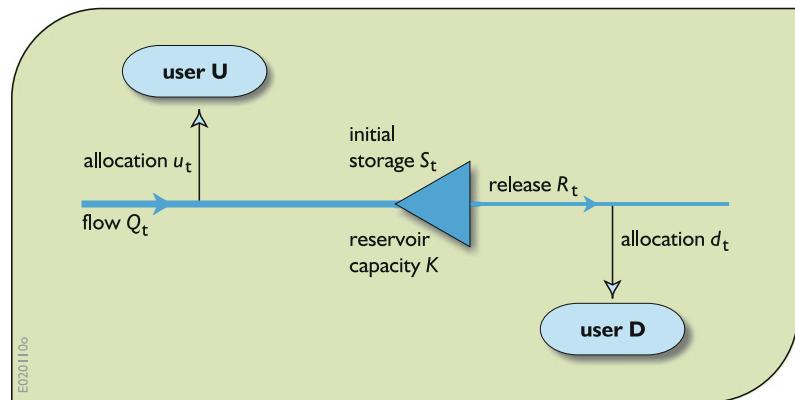
Note, when dividing a continuous known probability distribution into discrete ranges, the discrete flows q_i selected to represent each range i having a given probability PQ_i, should be selected so as to maintain at least the mean and variance of that known distribution as defined by Eqs. 7.5 and 7.6.

To illustrate this consider a slightly more involved 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 Fig. 7.10.

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 prespecified target allocations, each allocation in each time period will depend in part on the reservoir storage volume.

Consider first a deterministic model of the above problem, assuming known flows Q_t and upstream and downstream allocation targets UT_t and DT_t in each of T within-year periods t in a

Fig. 7.10 Example water resource system involving water diversions from a river both upstream and downstream of a reservoir of known capacity

year. Assume the objective is to minimize the sum of squared deviations of actual allocations, u_t and d_t, from their respective target allocations, UT_t and DT_t in each within-year period t.

$$\text{Minimize} \sum_{t}^{T} \left\{ (UT_t - u_t)^2 + (DT_t - d_t)^2 \right\} \tag{7.24}$$

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

$$S_t + Q_t - u_t - R_t = S_{t+1} \\ \text{for each period } t,\ T+1 = 1 \tag{7.25}$$

(b) Reservoir capacity limitations. Assuming a known active storage capacity K,

$$S_t \leq K \quad \text{for each period } t \tag{7.26}$$

Allocation restrictions for each period t:

$$u_t \leq Q_t \tag{7.27}$$

$$d_t \leq R_t \tag{7.28}$$

Equations 7.25 and 7.28 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 Eq. 7.25 would become an inequality.

The solution for this model, Eqs. 7.24–7.28, 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, as well flows, Q_t, as discussed in Chap. 4.

This deterministic discrete dynamic programming allocation and reservoir operation model

can be written for different discrete values of storage volumes S_t from $0 \le S_t \le$ capacity K as

$$F_t^n(S_t, Q_t) = \min\left\{(\mathrm{UT}_t - u_t)^2 + (\mathrm{DT}_t - d_t)^2 + F_{t+1}^{n-1}(S_{t+1}, Q_{t+1})\right\}$$

$$u_t, R_t, d_t$$
$$u_t \le Q_t$$
$$R_t \le S_t + Q_t - u_t$$
$$R_t \ge S_t + Q_t - u_t - K$$
$$d_t \le R_t$$
$$S_{t+1} = S_t + Q_t - u_t - R_t$$

(7.29)

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 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

$$\text{minimize}(\mathrm{UT}_t - u_t)^2 + (\mathrm{DT}_t - d_t)^2 \quad (7.30)$$

where

$$u_t \le Q_t \qquad (7.31)$$

$$d_t \le S_t + Q_t - u_t - S_{t+1} \qquad (7.32)$$

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 Eq. 7.29 can be written for all discrete storage volumes S_t from 0 to K and for all discrete flows Q_t as

$$F_t^n(S_t, Q_t) = \min_{\substack{S_{t+1} \\ S_{t+1} \le K}} \left\{ \begin{array}{l} (\mathrm{UT}_t - u_t(S_t, S_{t+1}))^2 + \\ (\mathrm{DT}_t - d_t(S_t, S_{t+1}))^2 \end{array} \right.$$

(7.33)

where the functions $u_t(S_t, S_{t+1})$ and $d_t(S_t, S_{t+1})$ have been determined using Eqs. 7.30–7.32.

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 likely 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, flow Q_t and period t. This constant is the optimal, in this case minimum, annual value of the objective function, Eq. 7.24.

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.
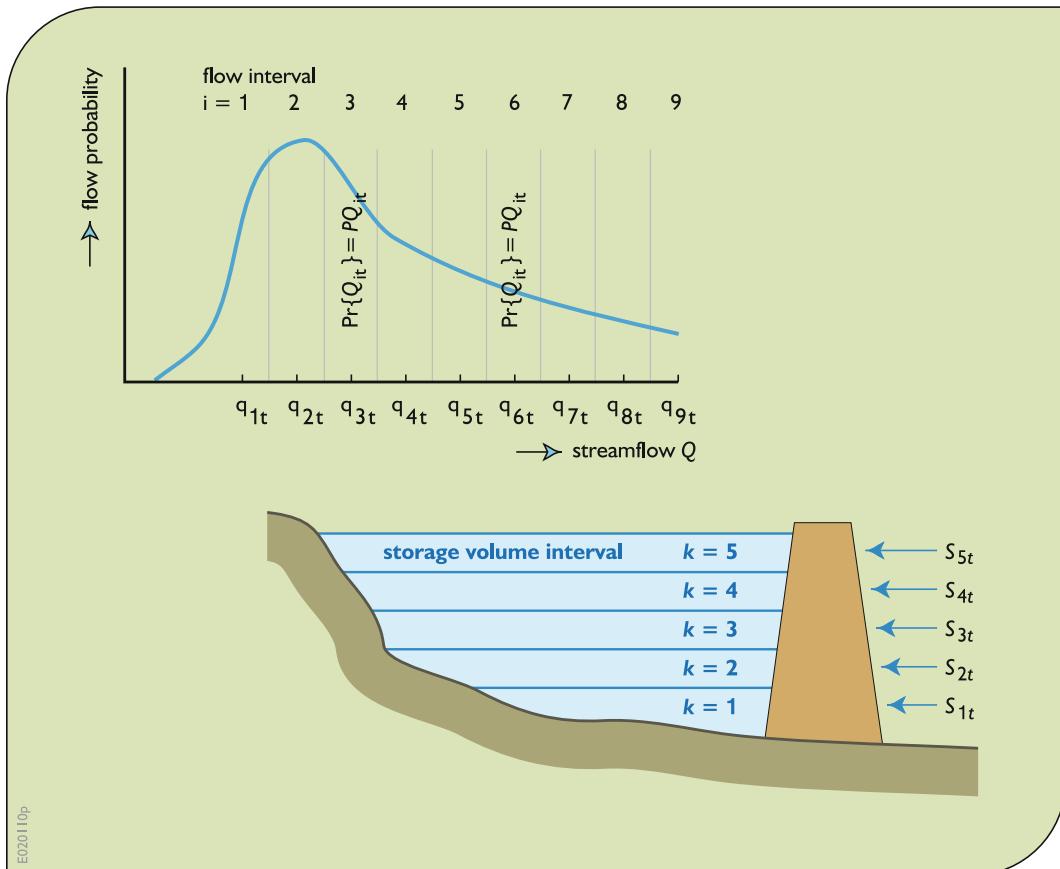
The above deterministic dynamic programming model (Eq. 7.33) can be converted to a stochastic model. Stochastic models consider multiple discrete flows as well as multiple discrete storage volumes, and their probabilities, in each period t. A common way to do this is to assume that the sequence of flows follow a first-order Markov process. Such a process involves the use of transition or conditional probabilities of flows as defined by Eq. 7.20.

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 index l denote 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 in 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 q_{it} and $q_{j,t+1}$ represent the flows in 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 Fig. 7.11.

Fig. 7.11 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}

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 Eqs. 7.30–7.32, and 7.33.

To identify the optimal allocations in each period t for each pair of feasible initial and final storage volumes S_{kt} and $S_{1,t+1}$, and inflows q_{it}, one can solve Eqs. 7.34–7.36.

$$\text{minimize}(\text{UT}_t - u_{kit})^2 + (\text{DT}_t - d_{kilt})^2 \quad (7.34)$$

where

$$u_{kit} \leq q_{it} \quad \text{for all } k, i, t. \quad (7.35)$$

$$d_{kilt} \leq S_{kt} + q_{it} - u_{kit} - S_{l,t+1} \\ \text{for all feasible } k, i, l, t. \quad (7.36)$$

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 7.33, 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

$$F_t^n(S_{kt}, q_{it}) = \min\{(\text{UT}_t - u_t(k,i))^2 + (\text{DT}_t - d_t(k,i,l))^2 \\ + \sum_j P_{ij}^t F_{t+1}^{n-1}(S_{l,t+1}, q_{j,t+1})\} \\ S_{l,t+1} \\ S_{l,t+1} \leq K \\ S_{l,t+1} \leq S_{kt} + q_{it}$$

$$(7.37)$$

Each P_{ij}^t 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.

$$P_{ij}^t = \text{Pr}\{\text{flow } q_{j,t+1}\text{within interval } j \text{ in } t+1|\text{flow of } q_{it} \\ \text{within interval } i \text{ in } t\}$$

The sum over all flow intervals j of these conditional probabilities times the $F_{t+1}^{n-1}(S_{l,t+1}, q_{j,t+1})$ 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 $F_t^n(S_{kt}, q_{it})$ is the expected minimum sum of squared deviations from the allocation targets with n periods remaining given an initial storage volume interval of S_{kt} and flow interval of q_{it}. Stochastic models such as these provide expected values of objective functions.

Another way to write the recursion equations of this model, Eq. 7.37, 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 $q_{j,t+1}$:

$$F_t^n(k, i) = \min\{(\text{UT}_t - u_t(k,i))^2 + (\text{DT}_t - d_t(k,i,l))^2 \\ + \sum_j P_{ij}^t F_{t+1}^{n-1}(l,j)\} \\ l \text{ such that} \\ S_{l,t+1} \leq K \\ S_{l,t+1} \leq S_{kt} + q_{it}$$

$$(7.38)$$

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 ℓ that identifies the best interval l given intervals k, i and period t.

$$l = \ell(k, i, t) \quad (7.39)$$

All values of l given k, i, and t, defined by Eq. 7.39, 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 $R_t(k, i)$.

$$S_{kt} + q_{it} - u_t(k,i) - R_t(k,i) = S_{l,t+1}$$
$$\forall k, i, t \text{ where } l = \ell(k,i,t) \qquad (7.40)$$

7.6.1   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.

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_{ij}^t = 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_{ij}^t can be displayed in matrices, similar to Table 7.3, 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

$$\sum_j P_{l,j,t+1} = \sum_k \sum_i P_{kit} \quad \forall l, t \qquad (7.41)$$

where the sums in the right hand side of Eq. 7.41 are over only those combinations of k and i that

result in a final volume interval l. This relationship is defined by Eq. 7.39 ($l = \ell(k, i, t)$).

While Eq. 7.41 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_{ij}^t 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.

$$P_{l,j,t+1} = \sum_k \sum_i P_{kit} P_{ij}^t \quad \forall l, j, t \quad l = \ell(k,i,t)$$
$$(7.42)$$

Once again the sums in Eq. 7.42 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.

$$\sum_k \sum_i P_{kit} = 1 \quad \forall t \qquad (7.43)$$

Note the similarity of Eqs. 7.42 and 7.43 to the Markov steady-state flow Eqs. 7.22 and 7.23. Instead of only one flow interval index considered in Eqs. 7.22 and 7.23, Eqs. 7.42 and 7.43 include two indices, one for storage volume intervals and the other for flow intervals. In both cases, one of Eqs. 7.22 and 7.42 can be omitted in each period t since it is redundant with that period's Eqs. 7.23 and 7.43, respectively.

The unconditional probabilities PS_{kt} and PQ_{it} can be derived from the joint probabilities P_{kit}.

$$PS_{kt} = \sum_i P_{kit} \quad \forall k, t \qquad (7.44)$$

$$PQ_{it} = \sum_k P_{kit} \quad \forall i, t \qquad (7.45)$$

Each of these unconditional joint or marginal probabilities, when summed over all their volume and flow indices, will equal 1. For example

$$\sum_k PS_{kt} = \sum_i PQ_{it} = 1 \qquad (7.46)$$

Note that these probabilities are determined based only on the relationships among flow and storage intervals as defined by Eq. 7.39, $l = \ell(k, i, t)$ in each period t, and the Markov chains defining the flow interval transition or conditional probabilities, P_{ij}^t. 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. (Amazing, isn't it?)

Given the values of these joint probabilities P_{kit}, the deterministic model defined by Eqs. 7.24–7.28 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.

$$\text{Minimize} \sum_k \sum_i \sum_t^T P_{kit} \left\{ (\text{UT}_t - u_{kit})^2 + (\text{DT}_t - d_{kit})^2 \right\}$$
$$(7.47)$$

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:

$$S_{kt} + q_{it} - u_{kit} - d_{kit} \geq S_{l,t+1} \quad \forall k, i, t$$
$$l = \ell(k, i, t) \qquad (7.48)$$

Reservoir capacity limitations.

$$S_{kit} \leq K \quad \forall k, i, t \qquad (7.49)$$

Allocation restrictions.

$$u_{kit} \leq q_{it} \quad \forall k, i, t \qquad (7.50)$$

7.6.2   A Numerical Example

A simple numerical example may help to illustrate how these stochastic models can be developed without getting buried in detail. Consider, for simplicity, 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 7.5.

Assuming equal within-year period durations, these three discrete flow rates are equivalent to

Table 7.5 Transition probabilities for two ranges of flows in two within-year periods

Q_j flow in $t = 2$		
Q_i flow in $t = 1$	3	6
1	0.6	0.4
3	0.3	0.7

Q_j flow in $t = 1$		
Q_i flow in $t = 2$	1	3
3	0.7	0.3
6	0.2	0.8

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 Eqs. 7.34–7.36 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 7.6 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 7.5 and 7.6, the dynamic programming model, Eq. 7.38 or as expressed in Eq. 7.51, 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 7.7.

$$F_t^n(k, i) = \min\left\{ \text{SD}_{kil} + \sum_j P_{ij}^t F_{t+1}^{n-1}(l,j) \right\}$$

$$\text{over all } l \text{ such that}$$

$$S_{l,t+1} \leq K$$

$$S_{l,t+1} \leq S_{kt} + Q_{it}$$

$$(7.51)$$

This process can continue until a steady-state policy is defined. Table 7.8 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 7.8 and 7.9.

Given this policy, the probabilities of being in any of these volume and flow intervals can be determined by solving Eqs. 7.42–7.45. Table 7.10 shows the results of these equations applied to the data in Tables 7.5 and 7.8. It is obvious that if the policy from Table 7.9 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 7.6 provides the annual expected squared deviations, associated with the selected discrete storage volumes and flows. This is done in Table 7.11 for those combinations of k, i, and l that are contained in the optimal solution as listed in Table 7.9.

The sum of products of the last two columns in Table 7.11 for each period t equals the expected squared deviations in the period. For period $t = 1$ the expected 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 nine iterations, ranging from 3475.5 to 3497.1 (as calculated from data in Table 7.8).

The policy for reservoir releases is a function not only of the initial storage volumes, but also of the current inflow, i.e., the total water available in the period. Reservoir release rule curves now must become two-dimensional. However, the inflow for each period usually cannot be predicted at the beginning of each period. Thus the reservoir release policy has to be

Table 7.6 Optimal allocations associated with given initial storage, S_k, flow, Q_i, and final storage, S_l, volumes

initial storage	flow	final storage	interval indices	upstream allocation	down-stream allocation	sum squared deviation
S_k	Q_i	S_1	k, i, l	u_{ki}	d_{kil}	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

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 $(S_k + Q_i - u_{ki} - S_l)$

Table 7.7 First four iterations of dynamic programming model, Eq. 7.51, moving backward in successive periods n, beginning in season $t = 2$ with $n = 1$

storage & flow k, i	period $t = 2, n = 1$		$SD_{kil} + \Sigma_j P_{ij}^t F_{t+1}^{n-1}(l, j)$	$F_t^n(k, i)$	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			

Table 7.7 (continued)

storage & flow k, i		period $t = 1, n = 2$ $SD_{kil} + \Sigma_j P_{ij}^t F_{t+1}^{n-1}(l, j)$	$F_t^n(k, i)$	optimal l
1,1	$l = 1$	$4996.0 + 0.6\,(1989.0) + 0.4\,(112.5)\quad = \mathbf{6234.4}$	6234.4	1
	$l = 2$	$7141.0 + 0.6\,(1152.0) + 0.4\,(\ \ 0.0)\quad = 7832.2$		
	$l = 3$	infeasible ---- $= ---$		
1,2	$l = 1$	$1989.0 + 0.3\,(1989.0) + 0.7\,(112.5)\quad = \mathbf{2664.45}$	2664.5	1
	$l = 2$	$3204.0 + 0.3\,(1152.0) + 0.7\,(\ \ 0.0)\quad = 3549.6$		
	$l = 3$	$4869.0 + 0.3\,(\ 544.5) + 0.7\,(\ \ 0.0)\quad = 5032.35$		
2,1	$l = 1$	$3301.0 + 0.6\,(1989.0) + 0.4\,(112.5)\quad = \mathbf{4539.4}$	4539.4	1
	$l = 2$	$4996.0 + 0.6\,(1152.0) + 0.4\,(\ \ 0.0)\quad = 5687.2$		
	$l = 3$	$7141.0 + 0.6\,(\ 544.5) + 0.4\,(\ \ 0.0)\quad = 7467.7$		
2,2	$l = 1$	$1152.0 + 0.3\,(1989.0) + 0.7\,(112.5)\quad = \mathbf{1827.45}$	1827.5	1
	$l = 2$	$1989.0 + 0.3\,(1152.0) + 0.7\,(\ \ 0.0)\quad = 2334.6$		
	$l = 3$	$3204.0 + 0.3\,(\ 544.5) + 0.7\,(\ \ 0.0)\quad = 3367.35$		
3,1	$l = 1$	$2056.0 + 0.6\,(1989.0) + 0.4\,(112.5)\quad = \mathbf{3294.4}$	3294.4	1
	$l = 2$	$3301.0 + 0.6\,(1152.0) + 0.4\,(\ \ 0.0)\quad = 3992.2$		
	$l = 3$	$4996.0 + 0.6\,(\ 544.5) + 0.4\,(\ \ 0.0)\quad = 5322.7$		
3,2	$l = 1$	$\ 544.5 + 0.3\,(1989.0) + 0.7\,(112.5)\quad = \mathbf{1219.95}$	1220.0	1
	$l = 2$	$1152.0 + 0.3\,(1152.0) + 0.7\,(\ \ 0.0)\quad = 1497.6$		
	$l = 3$	$1989.0 + 0.3\,(\ 544.5) + 0.7\,(\ \ 0.0)\quad = 2152.35$		

E020829d

(continued)

Table 7.7 (continued)

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	$l = 1$	$1989.0 + 0.7\,(6234.4) +0.3\,(2664.5) = 7152.4$		
	$l = 2$	$3204.0 + 0.7\,(4539.4) +0.3\,(1827.5) = \mathbf{6929.8}$	6929.8	2
	$l = 3$	$4869.0 + 0.7\,(3294.4) +0.3\,(1219.9) = 7541.1$		
1,3	$l = 1$	$112.5 + 0.2\,(6234.4) +0.8\,(2664.5) = 3490.0$		
	$l = 2$	$450.0 + 0.2\,(4539.4) +0.8\,(1827.5) = 2819.8$		
	$l = 3$	$1012.5 + 0.2\,(3294.4) +0.8\,(1219.9) = \mathbf{2647.3}$	2647.3	3
2,2	$l = 1$	$1152.0 + 0.7\,(6234.4) +0.3\,(2664.5) = 6315.4$		
	$l = 2$	$1989.0 + 0.7\,(4539.4) +0.3\,(1827.5) = \mathbf{5714.8}$	5714.8	2
	$l = 3$	$3204.0 + 0.7\,(3294.4) +0.3\,(1219.9) = 5876.1$		
2,3	$l = 1$	$0.0 + 0.2\,(6234.4) +0.8\,(2664.5) = 3378.4$		
	$l = 2$	$112.5 + 0.2\,(4539.4) +0.8\,(1827.5) = 2482.3$		
	$l = 3$	$450.0 + 0.2\,(3294.4) +0.8\,(1219.9) = \mathbf{2084.8}$	2084.8	3
3,2	$l = 1$	$544.5 + 0.7\,(6234.4) +0.3\,(2664.5) = 5707.9$		
	$l = 2$	$1152.0 + 0.7\,(4539.4) +0.3\,(1827.5) = 4877.8$		
	$l = 3$	$1989.0 + 0.7\,(3294.4) +0.3\,(1219.9) = \mathbf{4661.1}$	4661.1	3
3,3	$l = 1$	$0.0 + 0.2\,(6234.4) +0.8\,(2664.5) = 3378.4$		
	$l = 2$	$0.0 + 0.2\,(4539.4) +0.8\,(1827.5) = 2369.8$		
	$l = 3$	$112.5 + 0.2\,(3294.4) +0.8\,(1219.9) = \mathbf{1747.3}$	1747.3	3

E020829e

Table 7.7 (continued)

storage & flow k, i		period $t = 1, n = 4$ $\qquad SD_{kil} + \Sigma_j P_{ij}^t F_{t+1}^{n-1}(l,j)$	$F_t^n(k,i)$	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		

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

Table 7.8 Summary of objective function values $F_t^n(k, i)$ and optimal decisions for stages $n = 5$–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

Table 7.9 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

Table 7.10 Probabilities of flow and storage volume intervals associated with policy as defined in Table 7.9 for the example problem

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
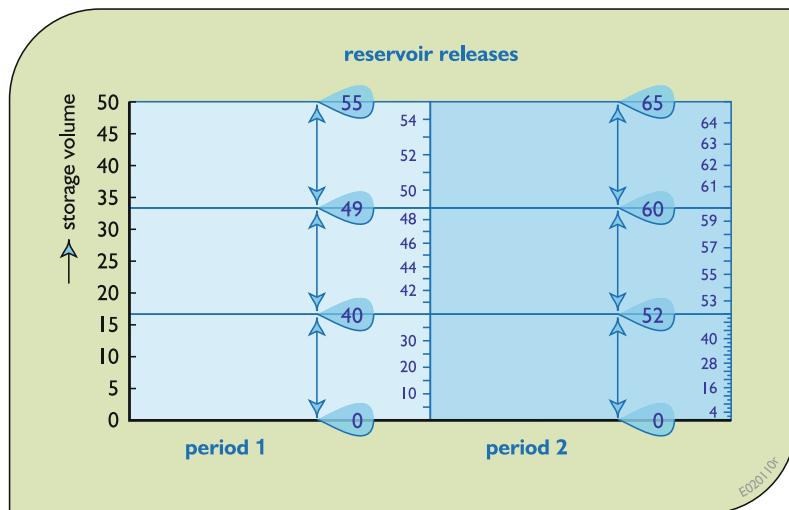
expressed in a way that it can be followed without knowledge of the current inflow. One way to do this is to compute the expected value of the release for each discrete storage volume, and show it in a release rule. This is done in Fig. 7.12. The probability of each discrete release associated with each 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 7.12 also shows the hedging that might take place as the reservoir storage volume decreases.

These and modifications of these policies can be simulated to determine improved release rules. Simulation modeling is the subject of the following chapter.

Table 7.11 The optimal operating policy and the probability of each state and decision

initial storage	flow	final storage	interval indices	time period	optimal allocation		sum squared deviations	joint probability
S_k	Q_i	S_1	k, i, l		u_{kit}	d_{kit}	SD_{kit}	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	2,3,2	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

Fig. 7.12 Reservoir release rule showing an interpolated release, increasing as storage volumes increase

7.7 Summary

This chapter has introduced some approaches for including risk into 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 be, 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 using 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 analyzed 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.

Reference

WWAP (United Nations World Water Assessment Programme) World Water Development Report 4. (2012). *Managing water under uncertainty and risk* (867 pp). Paris: UNESCO.

Additional References (Further Reading)

Bassson, M. S., Allen, R. B., Pegram, G. G. S., & van Rooyen, J. A. (1994). *Probabilistic management of water resource and hydropower systems*. Highlands Ranch, CO: Water Resources Publications.

Birge, J. R., & Louveaux, F. (1997). *Introduction to stochastic programming*. NY: Springer.

DeGroot, M. H. (1980). *Optimal statistical decisions*. NY: McGraw-Hill.

Dempster, M. A. H. (1980). *Stochastic programming*. London: Academic Press.

Dorfman, R., Jacoby, H. D., & Thomas, H. A., Jr. (1972). *Models for managing regional water quality*. Cambridge, MA: Harvard University Press.

Ermoliev, Y. (1970). *Methods of stochastic programming*. Moscow: Nauka. (in Russian).

Ermoliev, Y., & Wets, R. (Eds.). (1988). *Numerical techniques for stochastic optimization*. Berlin: Springer.

Heyman, D. P., & Sobel, M. J. (1984). *Stochastic models in operations research, Volume II. Stochastic optimization*. NY: McGraw-Hill.

Hillier, F. S., & Lieberman, G. J. (1990). *Introduction to stochastic models in operations research*. NY: McGraw-Hill.

Howard, R. A. (1960). *Dynamic programming and Markov processes*. Cambridge, MA: MIT Press.

Hufschmidt, M. M., & Fiering, M. B. (1966). *Simulation techniques for design of water-resource systems*. Cambridge, MA: Harvard University Press.

Loucks, D. P., Stedinger, J. S., & Haith, D. A. (1981). *Water resource systems planning and analysis*. Englewood Cliffs, NJ: Prentice Hall.

Maass, A., Hufschmidt, M. M., Dorfman, R., Thomas, H. A., Jr., Marglin, S. A., & Fair, G. M. (1962). *Design of water-resource systems*. Cambridge, MA: Harvard University Press.

Prékopa, A. (1995). *Stochastic programming*. Dordrecht, NL: Kluwer Academic Publishers.

Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory*. Cambridge, MA: Harvard University Press.

ReVelle, C. (1999). *Optimizing reservoir resources*. NY: Wiley.

Ross, S. M. (1983). *Introduction to stochastic dynamic programming*. NY: Academic Press.

Somlyódy, L., & Wets, R. J- B. Stochastic optimization models for lake eutrophication management. *Operations Research, 36*, 660–681.

Wets, R. J- B. (1989). Stochastic programming. In G. L. Nemhauser, A. H. G. Rinnooy Kan, & M. J. Todd (Eds.), *Handbooks in Operations Research and Management Science* (Vol. 1, pp. 573–629). North-Holland, Amsterdam: Elsevier Science Publishers.

Wurbs, R. A. (1996). *Modeling and analysis of reservoir system operations*. Upper Saddle River, NJ: Prentice Hall.

Exercises

7.1 Can you modify the deterministic discrete DP reservoir operating model to include the uncertainty, expressed as P_{ij}^t, of the inflows, as in Exercise 6.25?

(*Hints*: The operating policy would define the release (or final storage) in each season as a function of not only the initial storage but also the inflow. If the inflow changes, so might the release or final storage volume. Hence you need to discretize the inflows as well as the storage volumes. Both storage and inflow are state variables. Assume, for this model, you can predict with certainty the inflow in each period at the beginning of the period. So, each node of the network represents a known initial storage and inflow value. You cannot predict with certainty the following period's flows, only their probabilities. What does the network look like now?

7.2 Assume that there exist two possible discrete flows Q_{it} into a small reservoir in each of two periods t each year having probabilities P_{it}. Find the steady-state operating policy (release as a function of initial reservoir volumes and current period's inflow) for the reservoir that minimizes the expected sum of squared deviations from storage and release targets. Limit the storage volumes to integer values that vary from 3 to 5. Assume a storage volume target of 4 and a release target of 2 in each period t. (Assume only integer values of all states and decision variables and that each period's inflow is known at the beginning of the period.) Find the annual expected sum of squared deviations from the storage and release targets.

Period, t	Flows, Q_{it}		Probabilities, P_{it}	
	$i = 1$	$i = 2$	$i = 1$	$i = 2$
1	1	2	0.17	0.83
2	3	4	0.29	0.71

This is an application of Exercise 6.27 except the flow probabilities are independent of the previous flow.

7.3 Develop a linear model for defining the optimal joint probabilities of predefined discrete initial storage volumes, discrete inflows, and discrete final storage volumes in a reservoir in each period t. Let values of the index k represent the different discrete initial storage volumes, S_{kt}. Similarly, let the index i represent the inflows, Q_{it}, and the index l represent the final storage volumes, $S_{1,t+1}$, in period t. Let the index j represent the discrete inflows, $Q_{j,t+1}$, and m represent the discrete final storage volumes, $S_{m,t+2}$, in period $t + 1$. Let PR_{kilt} be the unknown joint probability of a discrete initial storage, S_{kt}, an inflow, Q_{it}, and a final storage volume, $S_{1,t+1}$, in period t. It is also the probability of a release associated with a particular combination of k, i, and l in period t. The objective is to maximize the expected net benefits, however, measured. The net benefits associated with any combination represented by k, i, and l in

period t is B_{kilt}. These net benefits and the conditional inflow probabilities. $P_{ij}^t = \Pr\{Q_{j,t+1}|Q_{it}\}$, are known. Show how the optimal operating policy can be determined once the values of the joint probabilities, PR_{kilt}, are known.

The same policy can be found by DP. Develop a DP model to find the optimal operating policy.

7.4 Referring to Exercise 7.3, instead of defining a final volume subscript l and m for computing joint probabilities PR_{kilt}, assume that subscripts d and e were used to denote different reservoir release volumes. How would the linear programming model developed be altered to include d and e in place of l and m? How would the dynamic programming recursion equation be altered?

7.5 Given joint probabilities PR_{kilt} found from Exercise 7.3, how would one derive the probability distribution of reservoir releases and storage volumes in each period t?

7.6 Assume that the streamflow Q at a particular site has cumulative distribution function $F_Q(q) = q/(1+q)$ for $q \geq 0$. The withdrawal x at that location must satisfy a chance constraint of the form $\Pr[x \geq Q] \leq 1 - \alpha$. Write the deterministic equivalent for each of the following chance constraints:
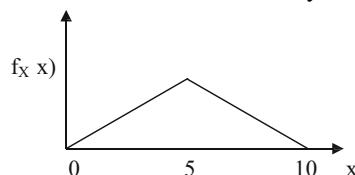
$$\Pr[x \leq Q] \geq 0.90 \quad \Pr[x \geq Q] \leq 0.80$$
$$\Pr[x \leq Q] \leq 0.95 \quad \Pr[x \leq Q] \leq 0.10$$
$$\Pr[x \geq Q] \geq 0.75$$

7.7 Monte Carlo Simulation:
Consider the symmetric triangular probability density function that ranges from 0 to 10 whose mean and most likely value is 5

(a) Generate values of x that come from this distribution.

To do this you need to

- Determine the equations of the cumulative distribution.
- Generate uniformly distributed random values of probabilities p.
- For each p find corresponding value of x. The inverse of the cumulative probability function $FX(x)$ denoted as $FX^{-1}(p)$.
- Using this inverse function, generate a series of 100 random variable values x that would have a probability distribution as shown above.

(b) Calculate the mean, variance, and standard deviation of this distribution based on the random values you computed. What is the effect on these statistics of increasing the number of sample values of the random variable, say from 100 to 1000 to 9000?

(c) Calculate and compare with the true mean and variance.

(d) Next, suppose these random values of X are flows entering a reservoir having a capacity of 6. The purpose of the reservoir is to release a target flow of 5 in each time period. Simulate the operation of the reservoir assuming that if there is insufficient water to meet the target release of 5, release what is available, leaving an empty reservoir. Find the mean, variance, and standard deviation of reservoir storage and release values.

(e) Finally assume the reservoir releases are to be allocated to three water users whose target allocations are 3, 2.33, and 8. Actual allocations should not exceed these target allocations. Make the allocations such that in each time period the maximum percentage deficit allocation is minimized. Find the mean, variance and standard deviation of each user's percentage deficit allocations.

System Sensitivity and Uncertainty Analysis

<div align="right">

8

</div>

The usefulness of any model is in part dependent on the accuracy and reliability of its output data. Yet, because all models are abstractions of reality, and because precise input data are rarely if ever available, all output values are subject to imprecision. The input data and modeling 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.

8.1 Introduction

Models are the primary way we have to estimate the multiple impacts of alternative water resource system design and operating policies. Models are used to estimate the values of various system performance indicators resulting from specific design and/or operating policy decisions. Model outputs are based on model structure, hydrologic and other time series inputs and a host of parameters whose values characterize 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 the model outputs or results will be wrong. Our models are always simplifications of the real systems we are analyzing. Furthermore, we simply cannot forecast the future with precision. So we know the model outputs defining future conditions are uncertain estimates, at best.

Some input data uncertainties can be reduced by additional research and further data collection and analysis. Before spending money and time to gather and analyze 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 if not eliminated. Such information helps determine how much one would be willing to "pay" to reduce model output uncertainty. If the uncertainty on average is costing a lot, it may pay to invest in additional data collection, in more studies, or in developing better models, all aimed at reducing that uncertainty. If that uncertainty 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 uncertainty is worthwhile, then the question is how best to do it. If doing this 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 benefits of more precise estimates of system performance, or the reduction of the uncertainty, that one can expect from obtaining such information. If additional information is to be obtained, it should be focused on that which reduces the uncertainties considered important, not the unimportant ones.

This chapter reviews some methods for identifying and communicating model output 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 analyses of large-scale regional systems. It builds on some of the statistical and stochastic modeling methods reviewed in the previous two chapters.

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 possible events or outcomes. Often risk refers to these probabilities times the magnitude of the consequences of these 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 estimated or described using probability distributions. 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 judgment, the risk is called subjective risk.
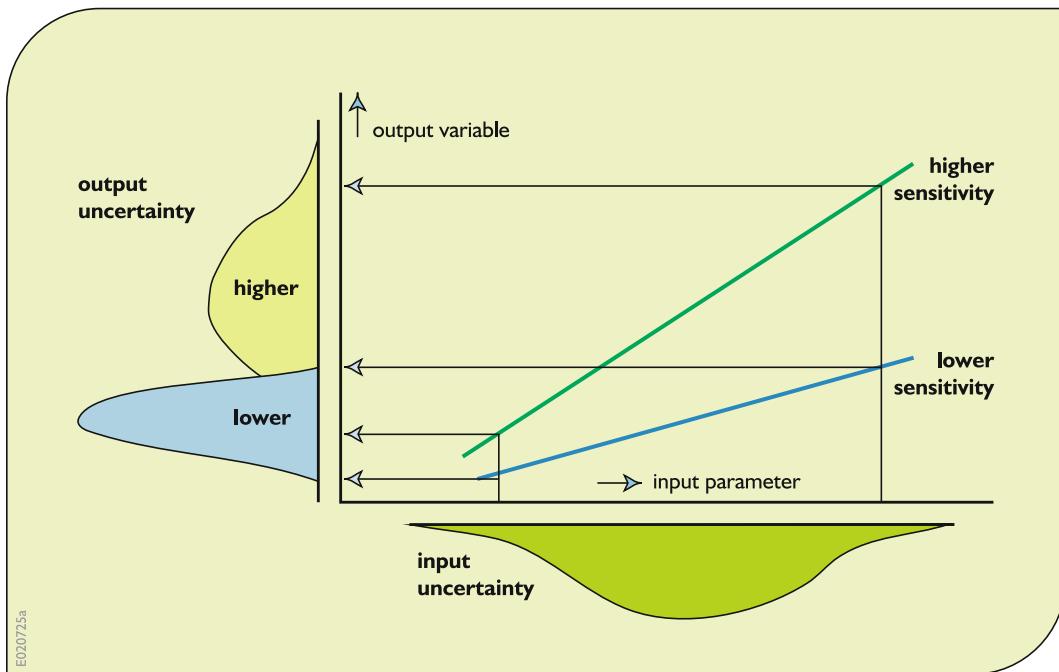
Such distinctions between objective and subjective risk, and between risk and uncertainty, rarely serve any useful purpose to those developing and using models. Likewise the distinctions are often unimportant to those who should be aware of the risks or uncertainties associated with system performance indicator values. If the probabilities associated with possible events or outcomes are unknown, and especially if the events themselves are unknown, then the approaches for performing sensitivity and uncertainty analyses will differ from those that are based on assumed known events and their probabilities.

Uncertainty in information is inherent in future-oriented planning efforts. Uncertainty stems from inadequate information and incorrect assumptions, as well as from the variability and possibly the nonstationarity of natural processes. Water managers often need to identify both the uncertainty as well as the sensitivity of system performance due to any changes in possible input data. They are often obligated to reduce any 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 changes (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 8.1 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 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 impacted ecosystems.

Fig. 8.1 Schematic diagram showing relationship among model input parameter uncertainty and sensitivity to model output variable uncertainty (Lal 1995)
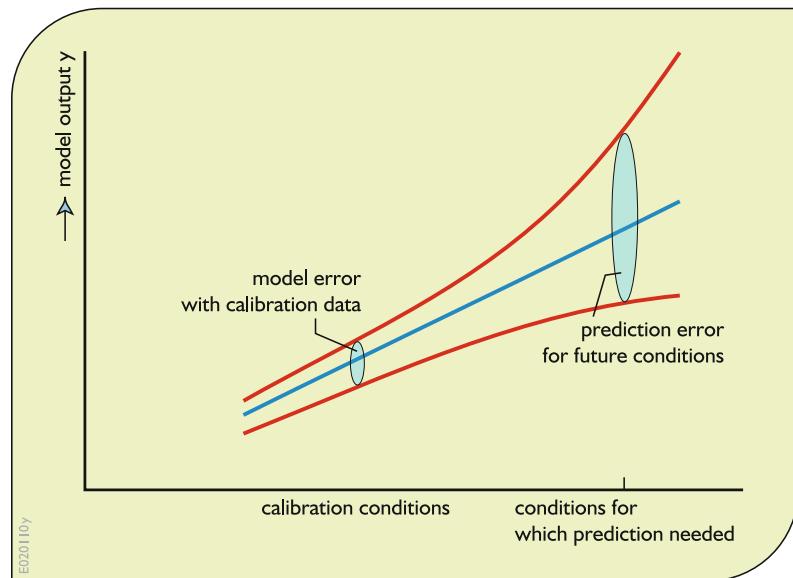
If asked how the system would operate with inputs similar to those observed in the past, the model should be able to provide a fairly precise estimate. Still that estimate will not be perfect. This is because our ability to reproduce current and recent operations is not perfect, though it should be fairly good. If asked to predict system performance for situations very different from those in the past, or when the historical data are not considered representative of what might happen in the future, say due to climate or technology change, such predictions become much less precise. There are two reasons why. First, our description of the characteristics of those different situations or conditions may be imprecise. Second, our knowledge base may not be sufficient for calibrating 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 are unlike those in the past, the less confidence we have that the model is providing a reliable description of systems operation. Figure 8.2 illustrates this issue.

Clearly a sensitivity analysis needs to consider how well a model can replicate current operations, and how similar the target conditions or scenarios are to those that existed in the past. 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 effect, but a small global effect. Changes in system-wide evapo-transpiration rates will likely impact system-wide flows. The precision of model projections and the relative importance of errors in different parameters will depend upon the:

(1) precision with which the model can reproduce observed conditions,
(2) difference between the conditions predicted in the future and the those that occurred in the past, and the
(3) system performance characteristics of interest.

Fig. 8.2 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

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. Some decisions may be able to be modified as new data and knowledge are obtained in a process of adaptive management.

There is also uncertainty with respect to human behavior and reaction related to particular outcomes and their likelihoods, i.e., 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 resource 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 favor actions that are robust (e.g., 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.

8.3 Variability and Uncertainty in Model Output

Differences between model output and observed values can result from either natural variability, say 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 much rainfall, evapotranspiration and consumption will occur, or just how a particular species will react to

various environmental and other habitat conditions. 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 8.3 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. 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 in the future and the relative importance of each of them.
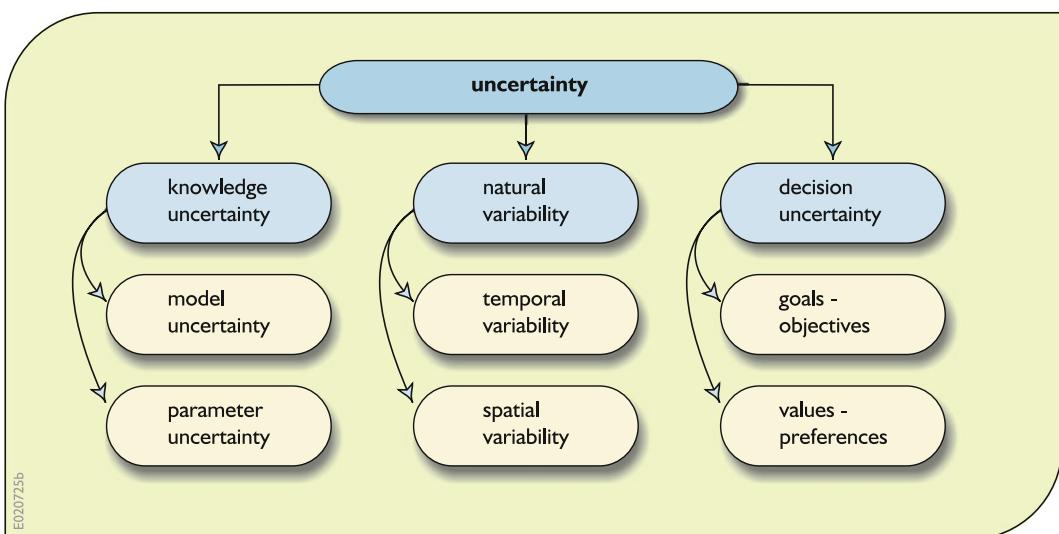
Rather than contrasting "knowledge" uncertainty versus natural variability versus decision uncertainty, one can classify uncertainty in another way based on specific sources of uncertainty, such as those listed below, and address ways of identifying and dealing with each source of uncertainty.

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

Fig. 8.3 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

8.3.1   Natural Variability

The main source of hydrologic 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 reason to think such events will continue to occur and become even more frequent and extreme. Research has demonstrated that climate has been variable in the past and concerns about anthropogenic activities that may increase that variability increase 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 in and on the border of the regions being studied.
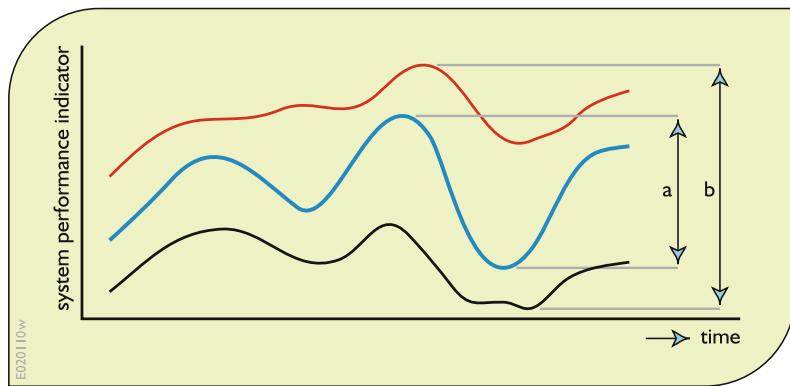
Time series input data are often actual, or at least based on, historical data. The time series values typically describe historical conditions including droughts and wet periods. What is distinctive about natural uncertainty, as opposed to errors and uncertainty due to modeling limitations, is that natural variability in meteorological forces cannot be reduced by improving the model's structure, increasing the resolution of the simulation, or by better calibration of model parameters.

Errors result if meteorological values are not measured or recorded accurately, or if mistakes are made when creating computer data files. Furthermore, there is no assurance the statistical properties of historical data will accurately represent the statistical properties 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 affect 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., hydrologic and meteorological data), the extent of the variability, and the lower and upper limits of that variability, may also be affected by errors in the inputs, the values of parameters, initial boundary conditions, model structure, processes and solution algorithms.
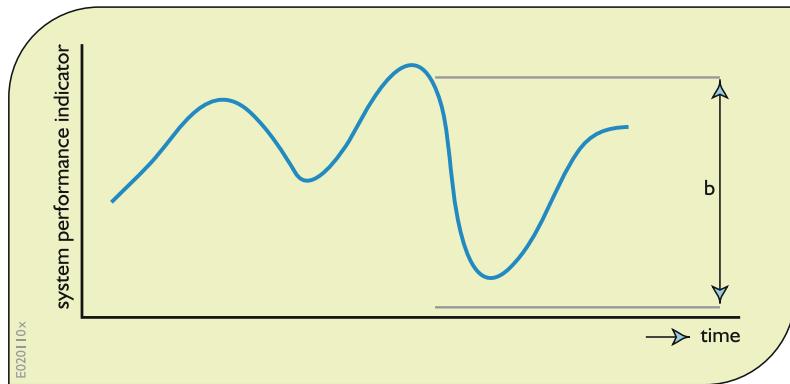
Figure 8.4 illustrates the distinction between the variability of a system performance indicator due to input data variability, and the extended range of variability due to the total uncertainty associated with any combination of the causes listed in the previous section. This extended range is what is of interest to water resource planners and managers.

In practice a time series of system performance indicator values 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 on the assumption that those measurements are representative of—come from the same statistical or stochastic process yielding—future measurements. Figure 8.5 illustrates this point. Note that the time series may even contain values outside

Fig. 8.4 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

Fig. 8.5 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

the range "*b*" defined in Fig. 8.4 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. Furthermore, it is unlikely one will ever know the 100% confident interval that includes all values that could ever occur.

8.3.2 Knowledge Uncertainty

Referring to Fig. 8.3, 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.

8.3.2.1 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 was repeated using different data sets, it would have been resulted in different parameter values. Those values would yield different simulated system behavior, 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 was 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 can be either fixed or variable. However, because they are not being computed based on 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.

8.3.2.2 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 to more closely represent the complexity of the real system may not only add to the cost of data collection, but 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 hydrologic modeling exercise with daily time steps where model uncertainty dominated parameter value uncertainty.

8.3.3 Decision Uncertainty

Uncertainty in model predictions can result from unanticipated changes in what is being modeled. 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? And can anyone identify a place where deviations from published policies do not occur? Policies implemented in practice tend to address short-term changes in policy objectives.

What humans will want to achieve in the future may not be the same as what they want today. Predictions of what people will want in the future are clearly sources of uncertainty. A perfect example of this is in the very flat Greater Everglades region of south Florida in the US. Sixty years ago, folks 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 a return to a more natural hydrologic system with more wetlands and unobstructed flows, but now for ecological restoration objectives that were not a major concern or much appreciated 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 changing social and economic processes influence human activities and their demands for water resources and environmental amenities over time.

Sensitivity scenarios that include human activities can help define the effects of those human activities within an area. It is important that these alternative scenarios realistically capture the forces or stresses that the system may face. The history of systems studies is full of examples where the issues studied were rapidly 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, sea-level rise, an act of terrorism, or even war. One thing is sure: the future will be different than the past, and no one can be certain just how.

8.3.3.1 Surprises

Water resource 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 some 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 are truly surprises. But surprises should be expected. Hence system flexibility and adaptability should be sought to deal with changing management demands, objectives, and constraints.

8.4 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 relative change in model output values given 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 various ways of performing and displaying sensitivity analyses are reviewed.
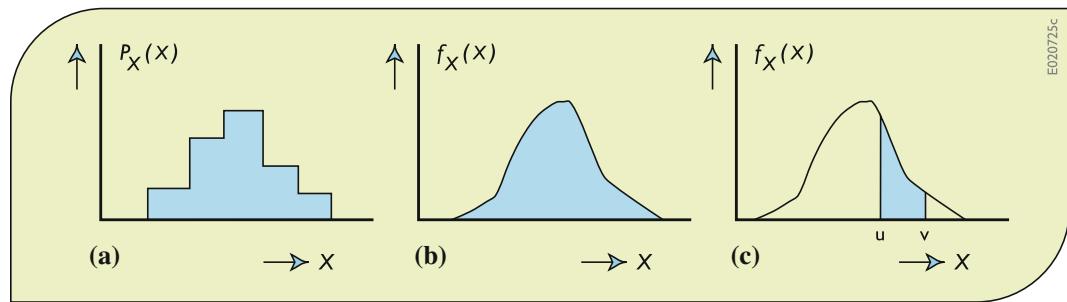
8.4.1 Uncertainty Analyses

Recall that uncertainty involves the notion of randomness. If a value of a performance indicator or performance measure, or in fact any variable (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. Here we are assuming we know, or can compute, or can estimate, this distribution.

Suppose the random variable is X. If the observed values of this random variable can be only discrete values, the probability distribution of X can be expressed as a histogram, as shown in Fig. 8.6a. 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 Fig. 8.6b. 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 Fig. 8.6c, represents the probability that the observed value x of the random variable value X will be within that range of values.

The probability distribution, $P_X(x)$ shown in Fig. 8.6a is called a probability mass function. The probability distributions shown in Fig. 8.6b, c are called probability density functions (pdf) and are denoted by $f_X(x)$. The subscript X of P_X and f_X represents the random variable, and the variable x (on the horizontal axes in Fig. 8.6) 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

Fig. 8.6 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

measures. Often targets associated with these indicators or measures are themselves uncertain.
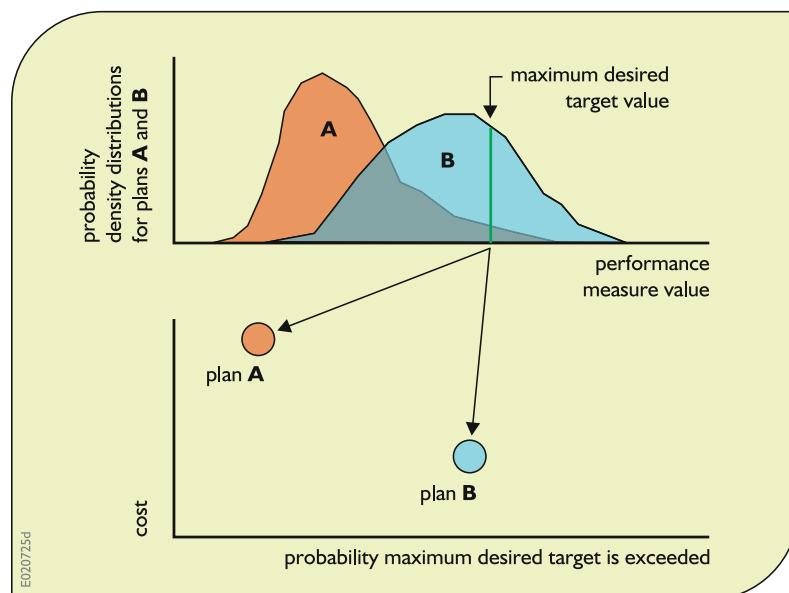
A complete uncertainty analysis would involve a comprehensive identification of all sources of uncertainty that contribute to the joint probability distributions of each input or output variable. Assume such analyses were performed for two alternative project plans, A and B, and that the resulting probability density distributions for a specified performance measure were as shown in Fig. 8.7. Figure 8.7 also identifies the costs of these two projects. The introduction of two performance criteria, cost and probability of
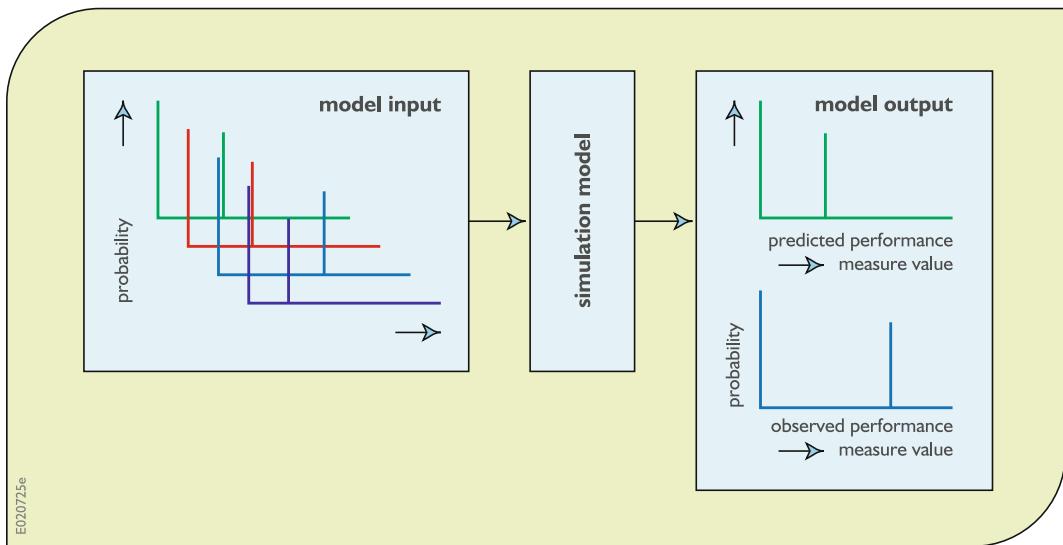
exceeding a performance measure target (e.g., a pollutant concentration standard) introduces a conflict where a tradeoff must be made.

8.4.1.1 Model and Model Parameter Uncertainties

Consider a situation as shown in Fig. 8.8, 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 randomly occurs. The model parameter values or model structure needs to be changed. This is typically done in a model calibration process.

Fig. 8.7 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)

Fig. 8.8 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 if found, corrected
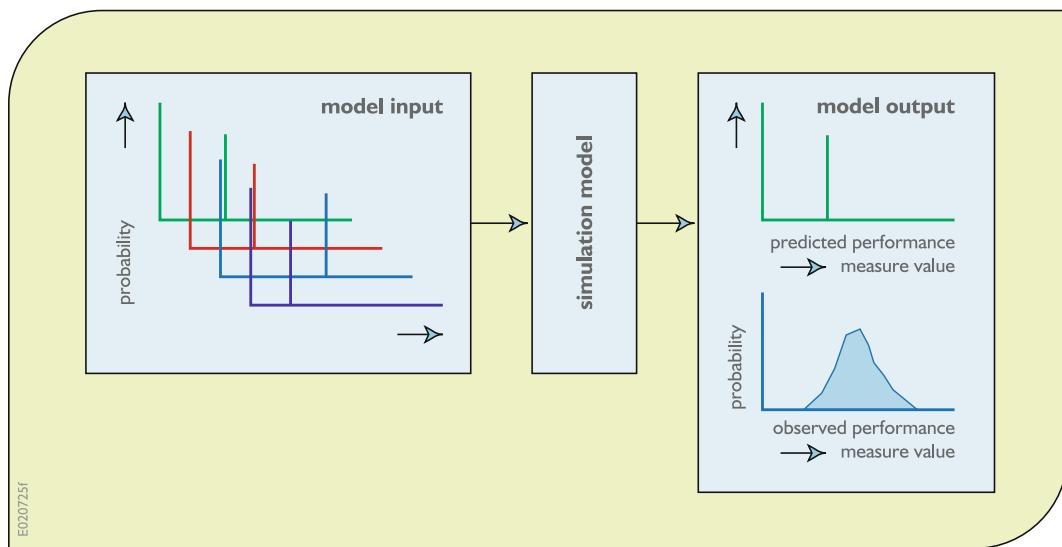
Given specific inputs, the outputs of deterministic models are always going to be the same each time those inputs are simulated. If for specified inputs to any simulation model the predicted output does not agree with the observed value, as shown in Fig. 8.8, this could result from imprecision in the 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.

Next consider the same deterministic simulation model but now assume at least some of the inputs are random, i.e., 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 Fig. 8.8. 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 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 the algorithms used to solve the models are correct and observed measurements of system performance vary for the same model inputs, as shown in Fig. 8.9, it can be assumed that the model structure does not capture all the processes that are taking place and that impact 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

Fig. 8.9 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

pollutant decay rate constant k (day^{-1}), and the pollutant discharge W (kg/day), we can use the following simplified 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 Fig. 8.9, 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 θ, and an additional model input, the streamflow temperature, T. It could be that the variation in streamflow temperature was the sole 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 nonpoint 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, the right equations are not known at the scale of application, or 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 taking this uncertainty into account. Imprecision will always exist due to less than a complete understanding of the system and the hydrologic processes being modeled. 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 model output uncertainty due to model structure.

The problem of determining the "optimal" level of modeling detail is particularly important when simulating the hydrologic 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 costless 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 hydrologic simulation models will typically involve considerable investments of money and time for data preparation and model calibration.
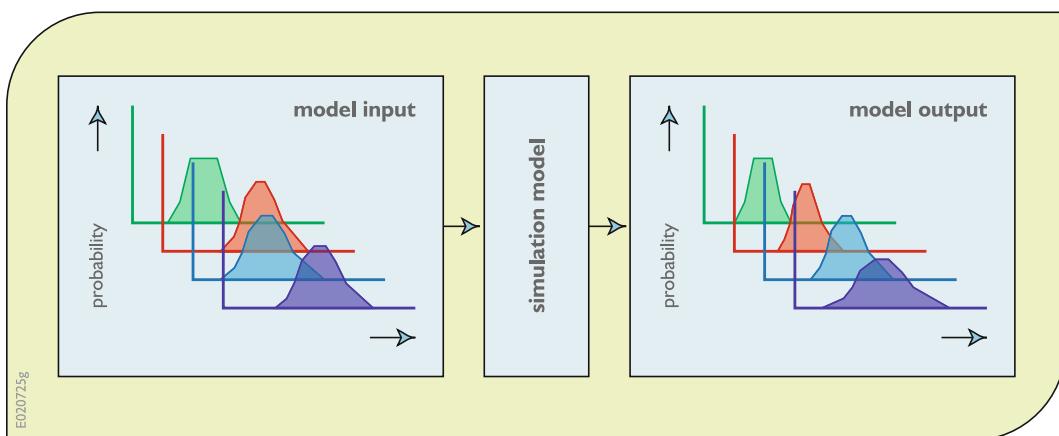
What is needed is a way to predict the variability evident in the system shown in Fig. 8.9. Instead of a fixed output vector for each fixed input vector, a distribution of outputs is needed for each performance measure based on fixed inputs (Fig. 8.9) or a distribution of inputs (Fig. 8.10). Furthermore, the model output distribution for each performance measure should "match" as well as possible the observed distribution of that performance measure.
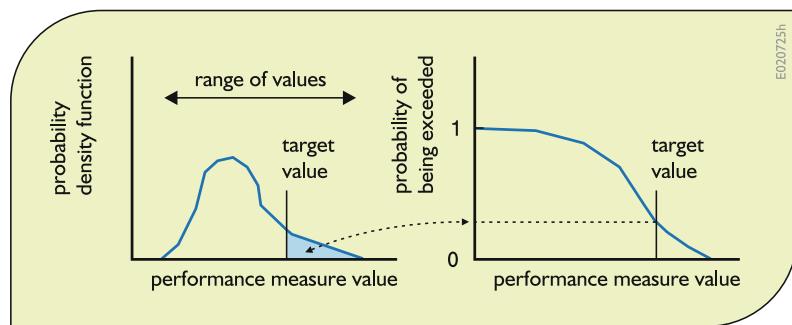
8.4.1.2 What Uncertainty Analysis Can Provide

An uncertainty analysis takes a set of randomly chosen input values (that 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 Fig. 8.11, 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.

Fig. 8.10 Simulating variable inputs to obtain probability distributions of predicted performance indices that match the probability distributions of observed performance values

Fig. 8.11 The distribution of performance measures defines 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

Common uses for uncertainty analyses are 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.
- Putting a reliability level on a function of the outputs, e.g., the range of function values that is likely to occur with some probability.
- Describing the likelihood of different potential outputs of the system.

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.

8.4.2   Sensitivity Analyses

"Sensitivity analysis" is aimed at describing 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 modeling process. The exact character of 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 watershed planners and managers, model users, and model 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 where efforts to improve models and their input values should be directed.

Initial sensitivity analysis studies could focus on two products:

(1) detailed results to guide research and assist model development efforts, and
(2) calculation of general descriptions of uncertainty associated with model predictions so that policy decisions can reflect both the predicted 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.

For 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 variable values. When such large uncertainties exist, predictions should be used with appropriate skepticism. Incremental strategies should be explored along with monitoring so that greater experience can accumulate to resolve some of those uncertainties.

Sensitivity analysis features are available in many linear and nonlinear 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 (Chap. 4). Thus sensitivity analysis addresses 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 should be of value to those making investment 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 impact 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 input data 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 outputs. 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.

8.4.2.1   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. The difficulty of

1. obtaining the derivatives for many models,
2. 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, and

3. having large variances associated with most hydrologic process models have motivated the replacement of analytical methods by numerical and statistical approaches to sensitivity analysis.

By varying the input probability distributions, one can determine the sensitivity of these distributions on the output distributions. If the output distributions 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_o. Let SI_{PQ} be the sensitivity index for an output variable Q with respect to a change ΔP in the value of the input variable P from its base value P_o. Noting that the value of the output $Q(P)$ is a function of P, a sensitivity index could be defined as

$$SI_{PQ} = [Q(P_o + \Delta P) - Q(P_o - \Delta P)]/2\Delta P \tag{8.1}$$

Other sensitivity indices could be defined (McCuen 1973). Let the index i represent a decrease and j represent an increase in the parameter value from its base value P_o, the sensitivity index SI_{PQ} for parameter P and output variable Q could be defined as

$$SI_{PQ} = \{|(Q_o - Q_i)/(P_o - P_i| + |(Q_o - Q_j)/(P_o - P_j)|\}/2 \tag{8.2}$$

or

$$SI_{PQ} = \max\{|(Q_o - Q_i)/(P_o - P_i)|, |(Q_o - Q_j)/(P_o - P_j)|\} \tag{8.3}$$

A dimensionless expression of sensitivity is the elasticity index, EI_{PQ} that measures the relative change in output Q for a relative change in input P.

$$EI_{PQ} = [P_o/Q(P_o)]SI_{PQ} \tag{8.4}$$

8.4.2.2 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, and observing the results.

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 and 95 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 8.1 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 ith 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

Table 8.1 Sensitivity of model output Y to possible errors in four parameter sets containing a single parameter or a group of parameters that vary together

parameter set		low value	nominal	high value
	1	$Y_{1,L}$	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

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 (milligrams per square meter per year, mg/m^2 a), the annual hydraulic loading, q (m/a or more exactly m^3/m^2 a), and the mean water depth, z (m).

$$P = (L/q)/\left[1 + (z/q)^{0.5}\right] \qquad (8.5)$$

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.

Data for Lake Ontario in North America would suggest that reasonable values of the parameters are $L = 680$ mg/m^{2}a; $q = 10.6$ 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 8.2.

One may want to display these results so they can be readily visualized and understood. A tornado diagram (Eschenback 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 having smallest limits on the bottom. Tornado diagrams (Fig. 8.12) are easy to construct and can include a large number of parameters without becoming crowded.

These error bars shown in Fig. 8.12 indicate there is substantial uncertainty associated with the phosphorus concentration P, primarily due to uncertainty in the loading rate L.

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 Fig. 8.13.
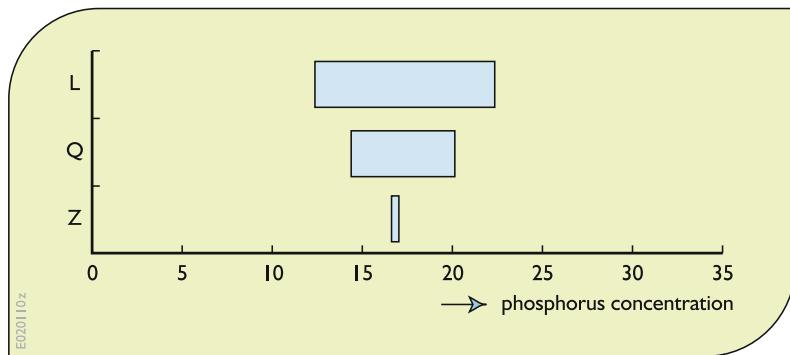
Another visual presentation is a spider plot showing the impact of uncertainty in each parameter on the variable in question, all on the same graph (Eschenback 1992; DeGarmo 1993, p. 401). A spider plot, Fig. 8.14, shows the

Table 8.2 Sensitivity of estimates of phosphorus concentration (mg/m^3) to model parameter values

		parameter value low	high	phosphorus concentration P_{low}	P_{high}
L —	P loading (mg/m². 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
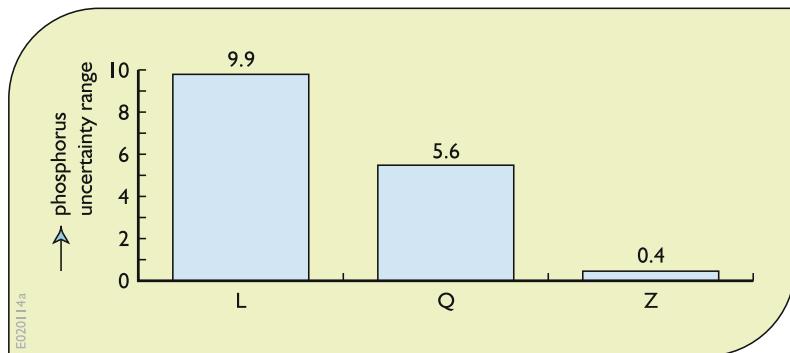
E020829n

The two right most values in each row correspond to the low and high values of the parameter, respectively
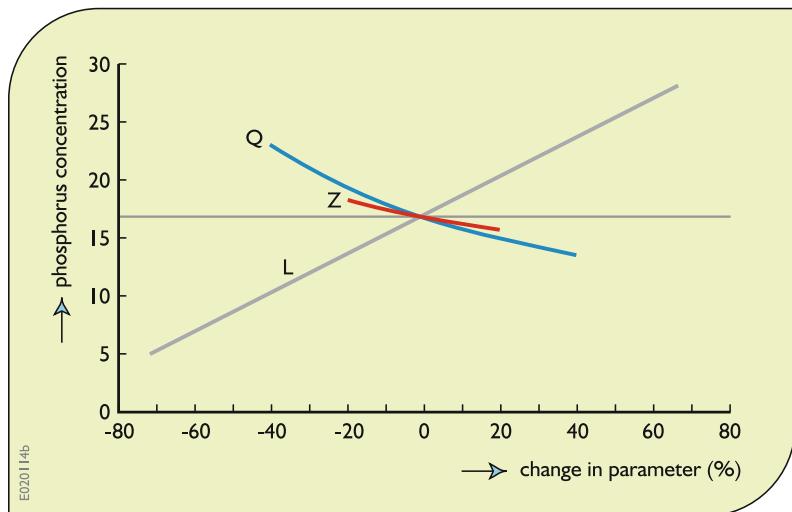
Fig. 8.12 A Tornado diagram showing the range of the output variable representing phosphorus concentrations for high and low values of each of the parameter sets. Parameters are sorted so that the largest range is on *top*, and the smallest on the *bottom*

Fig. 8.13 A Pareto Chart showing the range of the output variable representing phosphorus concentrations resulting from high and low values of each parameter set considered

Fig. 8.14 Spider Plot illustrates the relationships between phosphorus concentrations and variations in each of the parameter sets, expressed as a percentage deviation from their nominal values

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 4–5 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 nonlinear relationships 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 nonlinear relationship between P and q is illustrated. The range for z has been kept small given the limited uncertainty associated with that parameter.

8.4.2.3 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 example, possible errors in water depth would be accompanied with 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 modeled 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 three values (high, nominal, and low) 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 unlikely.

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 four 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.]
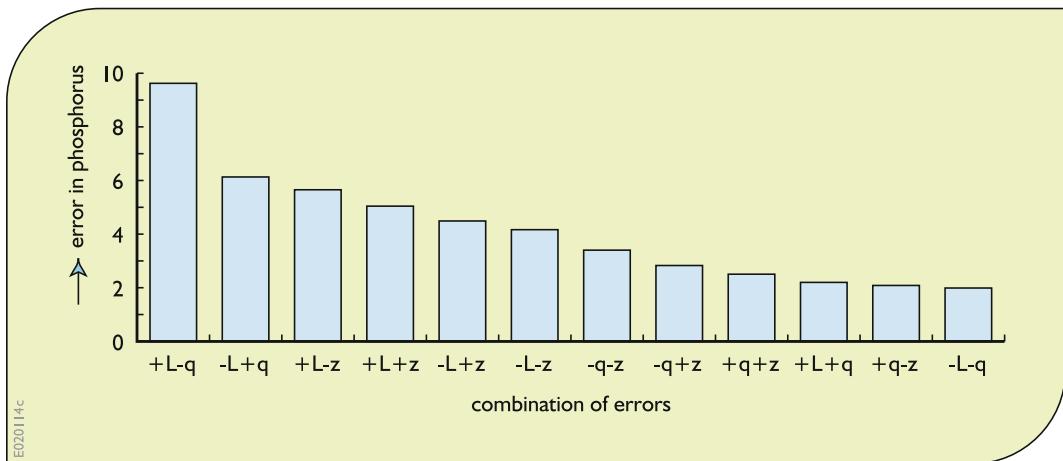
The presentation of these results could be simplified by displaying for each case only the maximum error, which would result in $m(m - 1)/2$ cases that might be displayed in a Pareto diagram. This would allow identification of those combinations of two parameters that might yield the largest errors and 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 Fig. 8.15.

Considering only the worst error for each pair of variables yields Fig. 8.16.

Here we see, as is no surprise, the worst error results from the most unfavorable combination of L and q values. If both parameters have their most unfavorable values, the predicted phosphorus concentration would be 27 mg/m^3.
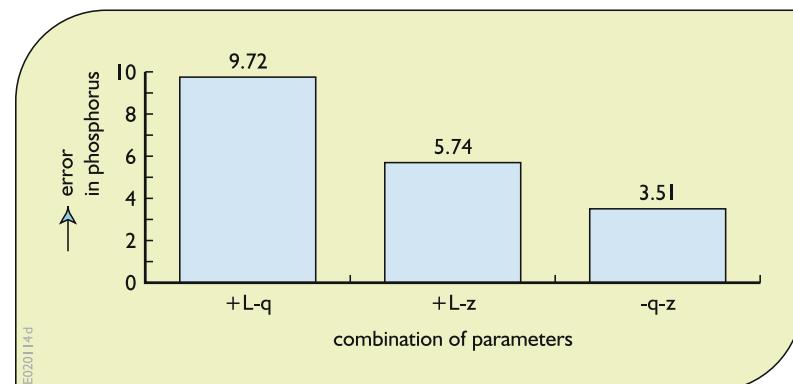
Looking for nonlinearities. One might also display in a Pareto diagram the maximum error for each pair as a percentage of the sum of the absolute values of the maximum error from each parameter separately. The ratio of the joint error to the individual errors would illustrate potentially important nonlinear interactions. If the

Fig. 8.15 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

Fig. 8.16 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

model of the system and the physical measure or economic metric were strictly linear, then the individual ratios should add to one.

8.4.2.4 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, that 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:

$$E[I] = F(\text{based on mean values of input parameters})$$
$$+ (1/2) \sum_i \sum_j [\partial F^2/\partial P_i \partial P_j] \text{ Cov } [P_i, P_j]\}$$

(8.6)

and

$$\text{Var}[I] = \sum_i \sum_j (\partial F/\partial P_i)(\partial F/\partial P_j) \text{Cov}[P_i, P_j]$$

(8.7)

where $(\partial F/\partial P_i)$ are the partial derivative 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.

$$\text{Cov}[P_i, P_j] = E[(P_i - E[P_i])(P_j - E[P_j])]$$

(8.8)

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

$$E[I] = F(\text{based on mean values of input parameters})$$

(8.9)

and

$$\text{Var } [I] = \sum_i [\partial F/\partial P_i]^2 \text{Var } [P_i]$$

(8.10)

Benjamin and Cornell (1970). Equation 8.6 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, p. 1523). This is true for any nonlinear function.

Of interest in the analysis of uncertainty is the approximation for the variance $\text{Var}[I]$ of indicator I. In Eq. 8.10 the contribution of P_i to the variance of I equals $\text{Var}[P_i]$ 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 8.3. 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.

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 and 95 percentiles is 1.645 standard deviations

Table 8.3 Approximate parameter sensitivity coefficients

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}	$[I_{3H}-I_{3L}]/[P_{3H}-P_{3L}]$

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

$$\mathrm{Var}[P_i] = \{[P_{i\mathrm{H}} - P_{i\mathrm{L}}]/3.3\}^2. \qquad (8.11)$$

This is all that is needed. Use of these average sensitivity coefficients is very reasonable for modeling the behavior 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 above. 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 8.4.

Assuming the parameter errors are independent:

$$\mathrm{Var}[P] = 9.18 + 2.92 + 0.02 = 12.12 \quad (8.12)$$

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 hydrologic loading q. Eliminating the uncertainty in z would have a negligible impact on the overall model error. Likewise, reducing the 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 ± 1.645 standard deviations around the mean define a 5–95 percentile interval, the 5–95 percentile interval would be about

$$16.8 \pm 1.645(3.48)\,\mathrm{mg/m}^3 = 16.8 \pm 5.7\,\mathrm{mg/m}^3$$
$$= 11.1 \text{ to } 22.5\,\mathrm{mg/m}^3.$$

$$(8.13)$$

The upper bound of 22.5 mg/m^3 is considerably less than the 27 mg/m^3 that would be obtained if both L and q had their most unfavorable 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 nonlinear systems with large uncertainty in

Table 8.4 Calculation of approximate parameter sensitivity coefficients

variable X	units	$\delta P/\delta X$	St Dev[X]	$(\delta P/\delta X)^2$ 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

the parameters. In such cases, Monte Carlo simulation (discussed below and in the previous chapter) or the use of higher order approximation may be required. Beck (1987, p. 1426) 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.

8.4.2.5 Fractional Factorial Design Method

An extension of first-order sensitivity analysis would be a more complete exploration 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. These two levels 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, 8 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 8.5 shows the 8 simulations required for a full factorial design method. The "+" and the "−" 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 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 jth simulation run. Then an estimate of the effect, denoted $\delta(Y|P_i)$ that input variable or parameter P_i has on the output

Table 8.5 A three-input factorial design

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

variable Y, is the average of the four separate effects of varying P_i:

For $i = 1$:

$$\delta(Y|P_1) = 0.25[(Y_2 - Y_1) + (Y_4 - Y_3)$$
$$+ (Y_6 - Y_5) + (Y_8 - Y_7)] \qquad (8.14)$$

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 this case, P_1 has no impact on the output variable Y.

Similarly the effects of P_2 and P_3, on variable Y can be estimated as:

$$\delta(Y|P_2) = 0.25\{(Y_3 - Y_1) + (Y_4 - Y_2)$$
$$+ (Y_7 - Y_5) + (Y_8 - Y_6)\} \qquad (8.15)$$

and

$$\delta(Y|P_3) = 0.25\{(Y_5 - Y_1) + (Y_6 - Y_2)$$
$$+ (Y_7 - Y_3) + (Y_8 - Y_4)\} \qquad (8.16)$$

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:

$$\delta(Y|P_1, P_2) = (1/2)\{[(Y_8 - Y_7) + (Y_4 - Y_3)]/2$$
$$- [(Y_2 - Y_1) + (Y_6 - Y_5)]/2\} = (1/4)\{[(Y_8 - Y_6)$$
$$+ (Y_4 - Y_2)] - [(Y_3 - Y_1) + (Y_7 - Y_5)]\}$$
$$(8.17)$$

Similar equations can be derived for looking at 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:

$$\delta(Y|P_3) = \delta(Y|P_1, P_2)$$
$$= 0.5\{(Y_8 - Y_3) - (Y_2 - Y_5)\} \quad (8.18)$$
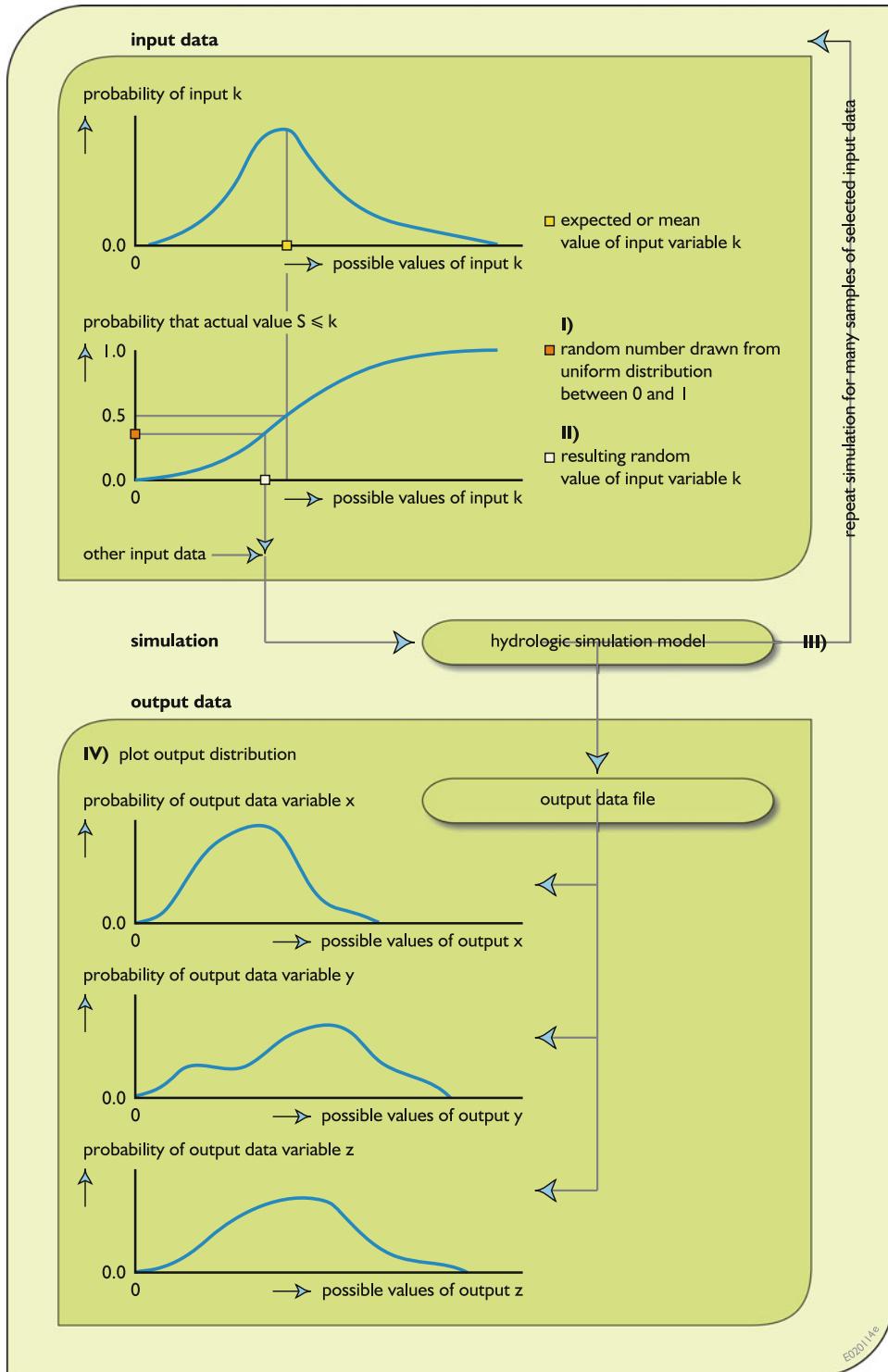
The separate effects of P_3 and of P_1P_2 are not available from the output. This is the loss in information resulting from fractional instead of complete factorial design.

8.4.2.6   Monte Carlo Sampling Methods

The Monte Carlo method of performing sensitivity analyses, illustrated in Fig. 8.17, 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 errors in 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 modeled system is simulated to obtain estimates of the selected performance indices. This must be done many times (often well over 100) to obtain a statistically significant description of system performance variability. The number of replications needed is generally not dependent on the number of parameters whose errors are to be analyzed. One can include in the simulation the 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 may result in unreasonable models. For example, model output based on calibrated 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

Fig. 8.17 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

Table 8.6   Monte Carlo analysis of lake phosphorus levels

parameter	L	q	z	P
specified means and standard deviations				
mean standard deviations	680.00	10.60	84.00	—
	121.21	1.67	1.82	---
generated means and standard deviations				
mean standard deviations	674.18	10.41	84.06	17.07
	130.25	1.73	1.82	3.61

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 based on these output variable values, is the computing time required. To reduce the computing times needed to perform sensitivity analyses using sampling methods, some tricks and as well as 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, Eq. 8.5, 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 8.6.
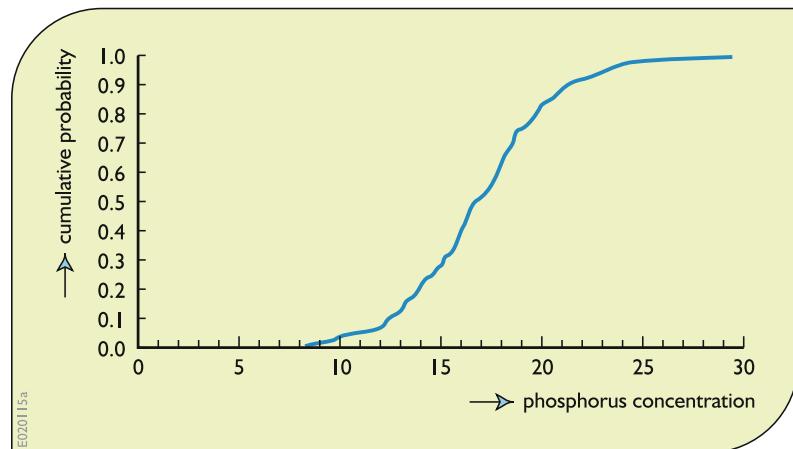
The table contains the specified means and variances for the generated values of L, q, and z, and also the actual values of the means and variances of the 200 generated values of L, q, z and also of the 200 corresponding generated output phosphorus concentrations, P. Figure 8.18 displays the distribution 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% to 95 percentile range was 11.1–23.4 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 different set of P values as well. Thus it is appropriate to estimate the

Fig. 8.18 Distribution of lake phosphorus concentrations from Monte Carlo analysis

standard error, SE, of this average. The standard error equals the standard deviation σ of the P values divided by the square root of the sample size n:

$$\text{SE} = \sigma/(n)^{0.5} = 3.61/(200)^{0.5} = 0.25. \quad (8.19)$$

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 16.6–17.6 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 8.7 shows, based upon the magnitudes of the correlation coefficients, errors in L were most important, and those in q second in importance.

One can also use regression to develop a linear model defining variations in the output based on errors in the various parameters. The results are shown in Table 8.8. The fit is very good, and $R^2 = 98\%$. If the model for P had been

Table 8.7 Correlation analysis of Monte Carlo results

variable	L	q	z	P
L	1			
q	0.079	1		
z	0.1297	-0.139	1	
P	0.851	-0.434	0.144	1

Table 8.8 Results of regression analysis on 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

linear, a R^2 value of 100% should have resulted. All of the coefficients are significantly different from zero.
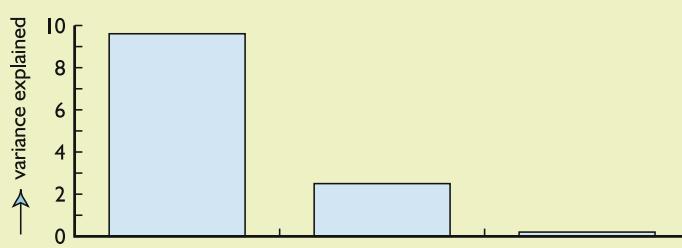
Note that the correlation between P and z was positive in Table 8.7, but the regression coefficient for z is negative. This occurred because there is a modest negative correlation between the generated z and q values. Use of partial correlation coefficients can also correct for such spurious correlations among input parameters.

Finally we display a plot, Fig. 8.19, based on this regression model illustrating the reduction in the variance of P that is due to dropping each variable 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 8.9.

Repeating the correlation analysis from before (shown in Table 8.10) now yields much clearer results that are in agreement with the regression analysis. The correlation between P and both

Fig. 8.19 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

Table 8.9 Standardized Monte Carlo analysis of lake phosphorus levels

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

Table 8.10 Correlation analysis of standardized Monte Carlo results

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

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.

Using control variates with a linear predictive model in conjunction with the standardized Monte Carlo variates, the standard deviation of the errors associated with the 200 observations is only 0.45. Thus the standard error for this estimate of the mean of P is $0.45/(200)^{0.5}$ or just 0.03. Thus this is a highly accurate result. The regressions were also repeated and yielded very similar results. The only real difference was that the parameter estimates had small standard errors and were more significant because of the elimination of correlation between the generated parameters.

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 which may reflect several dimensions and characteristics of model performance.

Let $L(P_i) > 0$ be the value of the likelihood measure assigned to the ith parameter set's calibration sequence. Then the model predictions generated with parameter set/combination P_i are assigned posterior probability, $p(P_i)$.

$$p(P_i) = L(P_i) / \sum_j L(P_j) \qquad (8.20)$$

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 centered in the best range (Beven 1993, p. 48). Then use of a different calibration period to generate the $p(P_i)$ 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, rather than the customary procedure of giving all the generated realizations equal weight. 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.

8.4.2.7 Latin Hypercube Sampling

For the simple Monte Carlo simulations described above, 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 a redundancy of information from repeated sampling in the same regions of a distribution and a lack of information from 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.
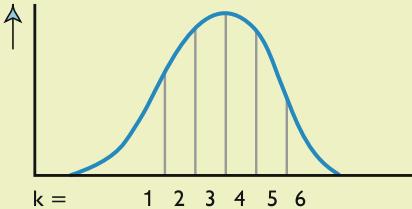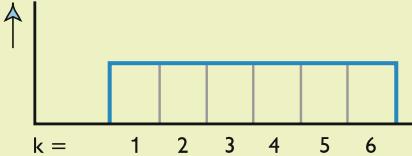
One such approach, Latin hypercube sampling (LHS), divides each input distribution into sections of equal probability for the specified

I)

randomly assign each of 6 intervals k
for each input j to 6 model-runs
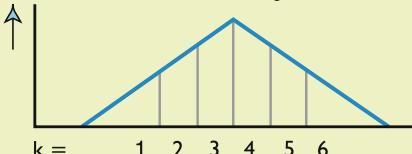
model run	randomly selected interval k for input P_j		
	P_1	P_2	P_3
1	4	3	5
2	2	1	6
3	1	6	1
4	6	4	4
5	3	2	3
6	5	5	2

normal distribution for P_1

k =   1 2 3 4 5 6

uniform distribution for P_2

k =   1   2   3   4   5   6

II)

randomly sample from each specified interval k
to obtain values P_{jk} for use in each simulation run
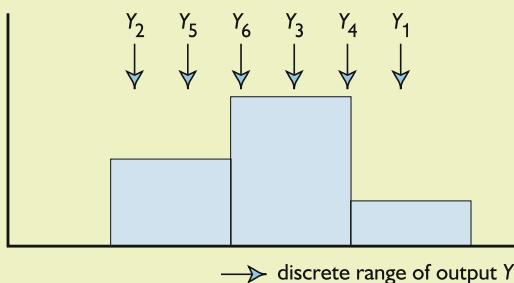
model run k	randomly selected values P_{jk} for each simulation run			value of output variable Y_k
	P_{1k}	P_{2k}	P_{3k}	
1	P_{14}	P_{23}	P_{35}	Y_1
2	P_{12}	P_{21}	P_{36}	Y_2
3	P_{11}	P_{26}	P_{31}	Y_3
4	P_{16}	P_{24}	P_{34}	Y_4
5	P_{13}	P_{22}	P_{33}	Y_5
6	P_{15}	P_{25}	P_{32}	Y_6

triangular distribution for P_3

k =   1   2   3   4   5   6

III)

plot histogram of output
resulting from model-runs

probability

Y_2 Y_5 Y_6 Y_3 Y_4 Y_1

discrete range of output Y
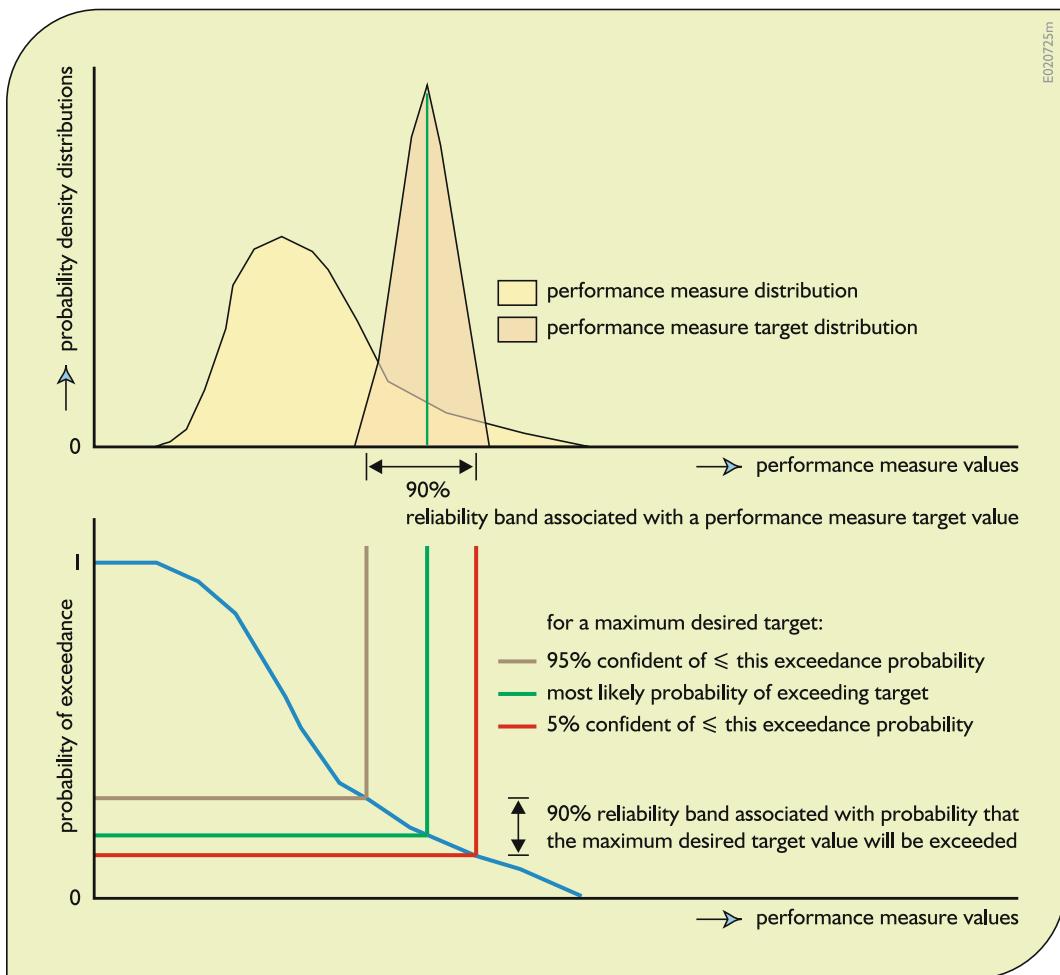
Fig. 8.20 Schematic representation of a Latin hypercube sampling procedure for six simulation runs

probability distribution, and draws one observation randomly from each range. Hence the ranges of input values within each section actually occur with equal frequency in the experiment. These values from each interval for each distribution are randomly assigned to those from other intervals to construct sets of input values for the simulation analysis. Figure 8.20 shows the steps in constructing a 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.

8.5 Performance Indicator Uncertainties

8.5.1 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

Fig. 8.21 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

are various ways of expressing the uncertainty associated with any target value. One such method is the use of qualitative approaches involving membership functions (Chap. 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 assume a distribution over a range of possible target values. Then this subjective probability distribution can be factored into the tradeoff analysis, as outlined in Fig. 8.21.

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 Fig. 8.20, suppose the target value represents some maximum limit of a pollutant, say phosphorus, concentration 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 uncertainty (probability of exceedance function) of pollutant concentrations—the performance measure—one can attach a confidence to any probability of exceeding the maximum desired concentration value.

The 95% probability of exceedance shown on Fig. 8.20, 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

Fig. 8.20 represents the most likely or median target value, the associated probability of exceedance is either the most likely, as indicated in Fig. 8.20, or that for which we are only 50% confident.
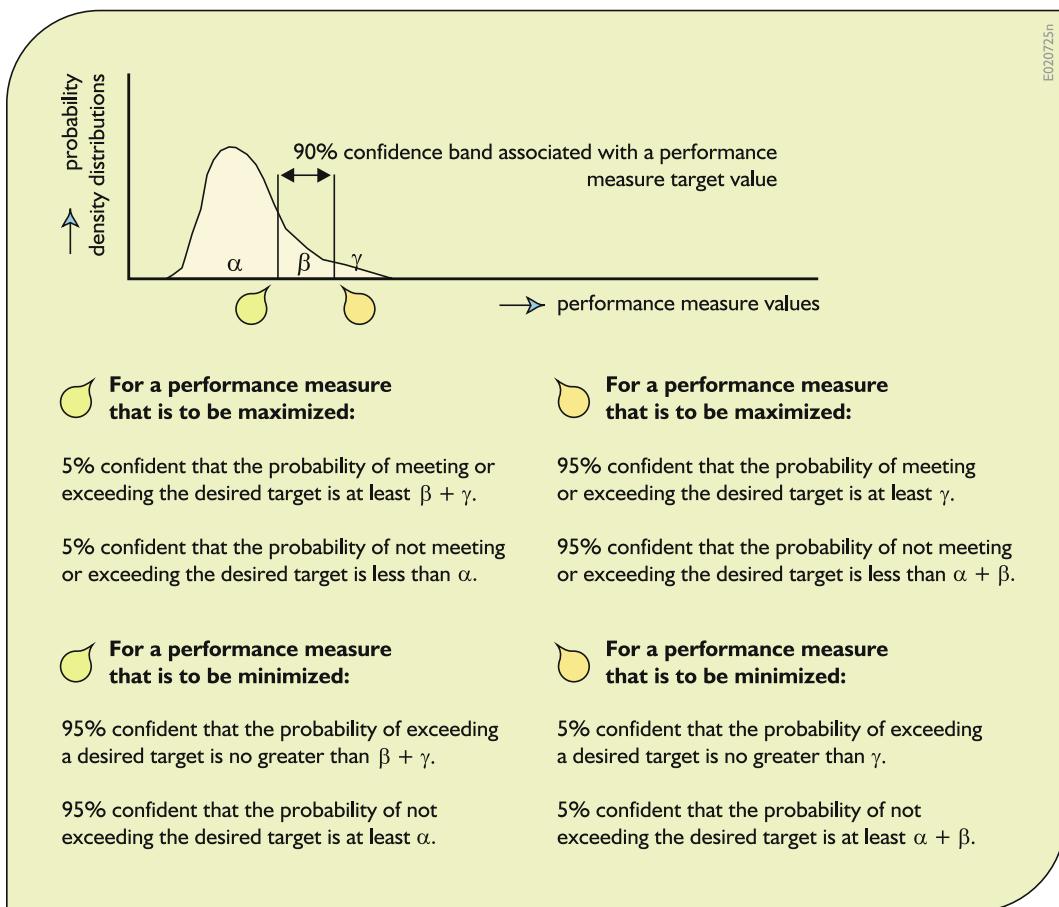
Figure 8.21 attempts to show how to interpret the reliabilities when the uncertain performance targets are

- 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 Fig. 8.22, one can state that one is 90% confident that the probability of achieving the desired target is no more than B. The 90% confidence level probability of not achieving the desired target is at least $A + C$. The probability of the performance measure being too low is at least A and the probability of the performance measure being too high is at least C, again at the 90% confidence levels. As the confidence 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

Fig. 8.22 Interpreting the results of combining performance measure probabilities with performance measure target probabilities depends on the type of performance measure. The letters *A*, *B*, and *C* represent proportions of the probability density function of performance measure values. (Hence probabilities *A* + *B* + *C* = 1)
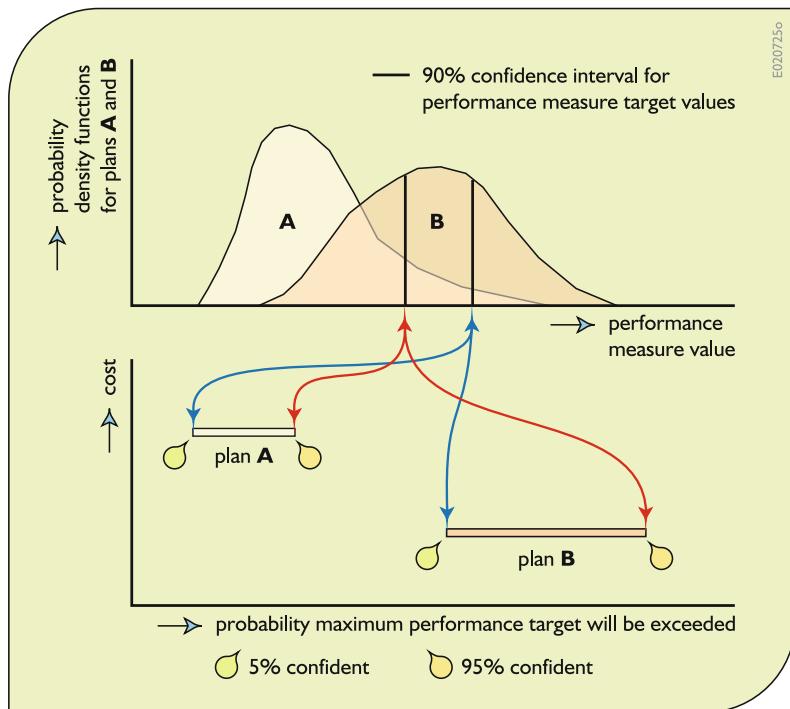
additional component of the plan in an evaluation process. Will plan evaluators and decision makers benefit from this additional information, and just how much additional uncertainty information is useful?

Now consider again the tradeoffs that need to be made as illustrated in Fig. 8.7. Instead of considering a single target value as shown on Fig. 8.7, assume there is a 90% confidence range associated with that single performance measure target value. Also assume that the target is a

maximum desired upper limit (e.g., of some pollutant concentration).
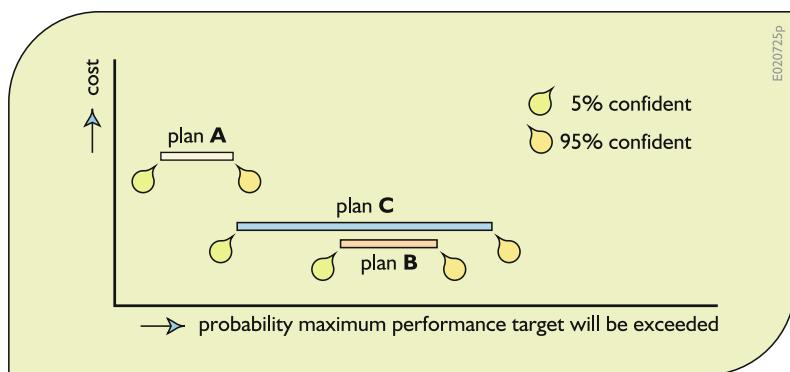
In the case shown in Fig. 8.23, the tradeoff is clearly between cost and reliability. In this example, no matter what confidence one chooses, Plan *A* is preferred to Plan *B* with respect to reliability, but Plan *B* is preferred to Plan *A* with respect to cost. The tradeoff is only between these two performance indicators or measures.

Consider however a third plan, as shown in Fig. 8.24. This situation adds to the complexity

Fig. 8.23 Two plans showing ranges of probabilities, depending on one's confidence, that an uncertain desired maximum (upper limit) performance target value will be exceeded. The 95% confidence levels are associated with the higher probabilities of exceeding the desired maximum target. The 5% confident 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*

Fig. 8.24 Tradeoffs among cost, reliabilities, and the confidence level of those reliabilities. The relative ranking of plans with respect to the probability of exceeding the desired (maximum limit) target may depend on the confidence given to that probability

of making appropriate tradeoffs. Now there are three criteria: cost, probability of exceedance (reliability) and the confidence in those reliabilities or 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 confidence in those probabilities.

In Fig. 8.23, in terms of cost the plans are ranked, from best to worst, B, C, and A. In terms of reliability at the 95% confidence level, they are ranked A, B, and C but at the 5% confidence 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 nor benefit in performing sensitivity and uncertainty analyses on all possible management alternatives. Rather one can focus on those alternatives that look the most promising, and then carry out additional uncertainty and sensitivity analyses only when important uncertain performance indicator values demand more scrutiny. Otherwise the work is not likely to affect the decision anyway.

8.5.2 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. Now the question is whether or not the observed differences are statistically significant. Can one really tell if one alternative is better than another or are the observed differences explainable by random variations attributable to variations in the inputs and how 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, is $E[Y_1] < E[Y_2]$? Alternatively one could check the difference in the median (50 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 nonparametric tests for paired and unpaired data can be employed.

Consider the differences between "paired" and "unpaired" data. Suppose that the meteorological data for 1941–1990 is used to drive a simulation model generating data as described in Table 8.11.

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, 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 1941–1970, and 1971–2000, they would have two sets of independent measurements for the

Table 8.11 Possible flow data from a 50-year simulation

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)

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: are the differences positive (say Y_1 tends to be larger then Y_2), or negative (Y_1 tends to be smaller), or are positive and negative differences are equally likely (there is no difference between Y_1 and Y_2).

Both parametric and nonparametric 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

$$X(i) = Y_1(i) - Y_2(i) \qquad (8.21)$$

is normally distributed. Then the hypothesis of no difference is rejected if the T statistic is sufficiently large, given the sample size n.

Alternatively, one can employ a nonparametric 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 nonparametric 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 nonparametric Wilcoxon Signed Rank test require that the differences between the simulated values for each year be computed.

8.6 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 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 favorably 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 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 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 illustrated in Figs. 8.10 and 8.27 be identified? Quantitative uncertainty analysis often can be used as the basis of qualitative information about uncertainty, even if the quantitative information is not what is communicated to the public.
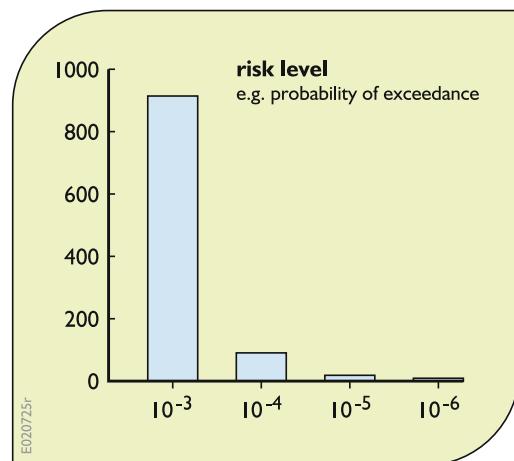
One should acknowledge to the public the widespread confusion regarding the differences between variability and uncertainty. Variability does not change through further measurement or study, although better sampling can improve our

knowledge about variability. Uncertainty reflects gaps in information about scientifically observable phenomena.

While it is important to communicate uncertainties and confidence in predictions, it is equally important to clarify who or what is at risk, 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 confidence associated with such predictions. A two-way dialog between those receiving such information and those giving it can help identify just what seems best for a particular audience.

Risk and uncertainty communication is a two-way street. It 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. Citizen advisory panels can give planners and decision-makers a better understanding of the questions and concerns of the community and an opportunity to test its effectiveness in communicating concepts and specific issues regarding uncertainty.
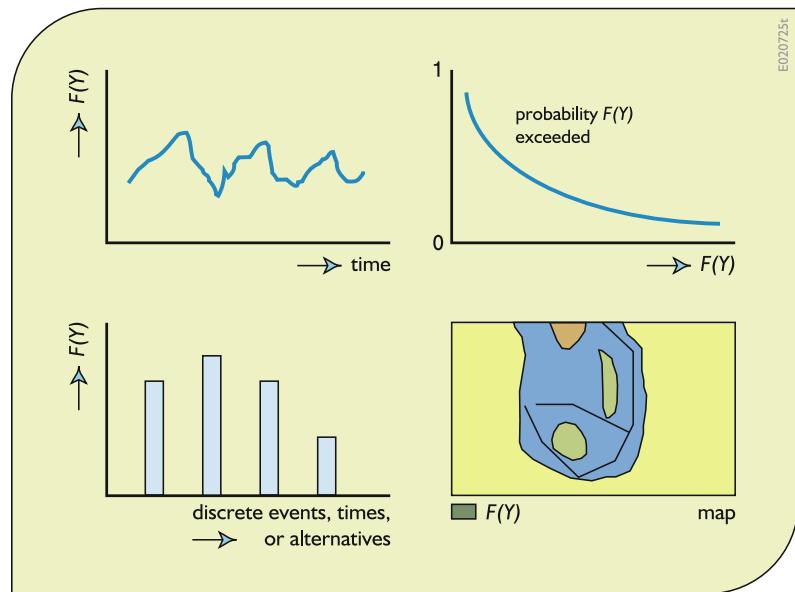
Fig. 8.25 Reducing risk by orders of magnitude is not equivalent to linear reductions

One approach to make uncertainty more meaningful is to make risk comparisons. For example, a ten-parts-per-billion target for a particular pollutant concentration is equivalent to 10 s in over 31 years. If this is an average daily concentration target that is to be satisfied "99%," of the time, 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. An alternative way to illustrate orders of magnitude of risk reduction is shown in Fig. 8.25, in which a bar graph depicts better than words that a reduction in risk from one in a 1000 (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 10-fold less than the first reduction of 90%. The percent of the risk that is reduced by whatever measures is an 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 minimizing a risk (NRC

Fig. 8.26 Different types of displays used to show model output Y or system performance indicator values $F(Y)$
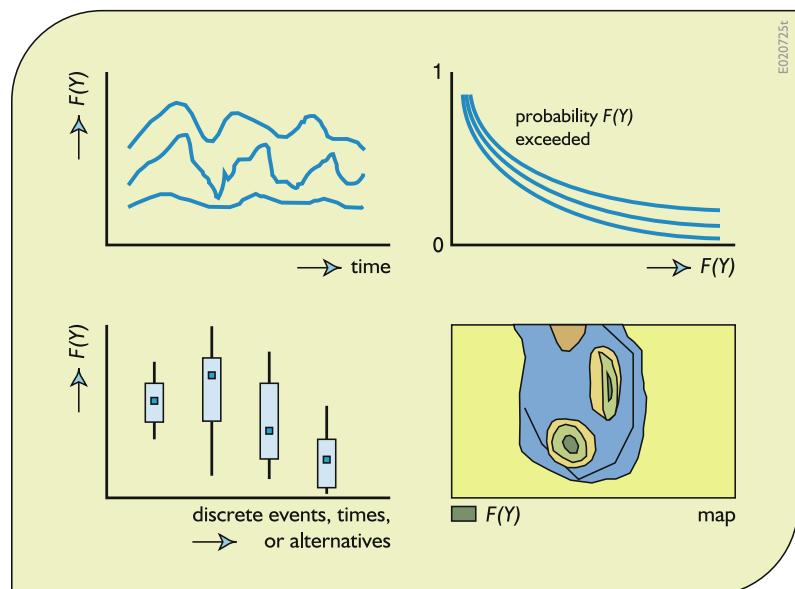
1989). One difficulty in using risk comparisons is that it is not always easy to find risks that are sufficiently similar to make a comparison meaningful. How is someone able to compare two alternatives having two different costs and two different risk levels, for example, as is shown in Fig. 8.7? One way is to perform an indifference analysis (as discussed in the next chapter), but that can lead to different 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 preferred tradeoff may differ.

At a minimum, graphical displays of uncertainty can be helpful. Consider the common system performance indicators that include:

Fig. 8.27 Plots of ranges of possible model output Y or system indicator values $F(Y)$ for different types of displays

- Time series plots for continuous time-dependent indicators (Fig. 8.26 upper left)
- Probability exceedance distributions for continuous indicators (Fig. 8.26 upper right),
- Histograms for discrete event indicators (Fig. 8.26 lower left), and
- Overlays on maps for space-dependent discrete events (Fig. 8.26 lower right).

The first three graphs in Fig. 8.26 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 interval. For overlays of information on maps, different colors could represent the spatial extents of events associated with different ranges of risk or uncertainty. Figure 8.27, corresponding to Fig. 8.26, illustrates these approaches for displaying these ranges.

8.7   Conclusions

This chapter provides an overview of uncertainty and sensitivity analyses in the context of hydrologic or water resources systems simulation modeling. 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 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, Monte Carlo methods for uncertainty analyses may be a very major and unattractive undertaking. Therefore it is often prudent to begin with the relatively simple deterministic procedures. This 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 decision 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 analyze 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 impacted by those decisions. Meeting those needs should result in more informed decision-making. But it comes at a cost that should be considered along with the benefits of having this sensitivity and uncertainty information.

References

Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water Resources Research, 23*(8), 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, NY.

Benjamin, J. R., & Cornell, C. A. (1970). *Probability, statistics and decision for civil engineers*. New York: McGraw-Hill Book Co.

Beven, K. (1989). Changing ideas in hydrology—The case of physically-based models. *Journal of Hydrology, 105*, 157–172.

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrologic modelling. *Advances in Water Resources, 16*, 41–51.

Binley, A. M., & Beven, K. J. (1991). Physically-based modelling of catchment hydrology: A likelihood approach to reducing predictive uncertainty. In D. G. Farmer & M. J. Rycroft (Eds.), *Computer modelling in the environmental sciences*. Oxford: Clarendon Press.

DeGarmo, E. P., Sullivan, W. G., & Bontadelli, J. A. (1993). *Engineering economy*. New York: MacMillian.

Devore, J. (1991). *Probability and statistics for engineering and the sciences* (3rd ed.). London: Brooks/Cole Publishing Co.

Eschenback, T. G. (1992). Spider plots versus tornado diagrams for sensitivity analysis. *Interfaces, 22*, 40–46.

Gaven, D. C., & Burges, S. J. (1981). Approximate error bounds for simulated hydrograph. *Journal of the Hydraulics Division, 107*(11), 1519–1534.

Kuczera, G. (1988). On the validity of first-order prediction limits for conceptual hydrologic models. *Journal of Hydrology, 103*, 229–247.

Lal, W. (1995, November). *Sensitivity and uncertainty analysis of a regional model for the natural system of South Florida*. South Florida Water Management District, West Palm Beach, FL, Draft report.

McCuen, R. (1973). The role of sensitivity analysis in hydrologic modeling. *Journal of Hydrology, 18*(1), 37–53.

National Research Council (NRC). (1989). *Improving risk communication*. Washington, D.C.: National Academy Press.

National Research Council (NRC). (1994). *Science and judgment in risk assessment*. Washington, D.C.: National Academy Press.

Stedinger, J. R., & Taylor, M. R. (1982). Synthetic streamflow generation, part II. *Effect of Parameter Uncertainty, Water Resources Research, 18*(4), 919–924.

Vollenweider, R. A. (1976). Advances in defining critical loading levels for phosphorus in lake eutrophication. *Memorie dell'Istituto Italiano di Idrobiologia, 33*, 53–83.

Additional References (Further Reading)

Anderson, J. L. (1998). Embracing uncertainty: The interface of Bayesian statistics and cognitive psychology. *Conservation Ecology, 2*(1), 2. Available from http://www.consecol.org/vol2/iss1/art2

Berry, D. A. (1996). *Statistics: A Bayesian perspective*. Belmont, CA: Duxbury.

Chapra, S. C., & Reckhow, K. H. (1979). Expressing the phosphorus loading concept in probabilistic terms. *Journal of the Fisheries Board of Canada, 36*, 225.

Colwell, R. K. (1974). Predictability, consistency, and contingency of periodic phenomena. *Ecology, 55*, 1148–1153.

Covello, V. T. (1987). Decision analysis and risk management decision-making: Issues and methods. *Risk Analysis, 7*, 131–139.

Covello, V. T., & Merkhofer, M. W. (1993). *Risk assessment methods*. London: Plenum Press.

Deutsch, C. V., & Journel, A. G. (1992). GS-LIB: Geostatistical software library and user's guide. Oxford: Oxford University Press.

Devore, J. L., & Peck, R. (1996). *Statistics: The exploration and analysis of data* (3rd ed.). London: Brooks/Cole Publishing Co.

Dilks, D. W., Canale, R. P., & Meier, P. G. (1992). Development of Bayesian Monte Carlo techniques for water quality model uncertainty. *Ecological Modelling, 62*, 149–162.

Dilks, D. W., & James, R. T. (2002). Application of Bayesian Monte Carlo analysis to determine the uncertainty in the Lake Okeechobee water quality model. In *Proceedings of Water Environment Federation, Watershed*.

DOE. (1998). Screening assessment and requirements for a comprehensive assessment, DOE/RL-96-16, Rev. 1, U.S. Department of Energy, Richland, Washington.

Duan, Q., Soroosian, S., & Ibbitt, R. P. (1988). A maximum likelihood criterion for use with data collected at unequal time intervals. *Water Resources Research, 24*(7), 1163–1173.

Fedra, K. (1983). A Monte Carlo approach to estimation and prediction. In M. B. Beck & G. Van Straten (Eds.), *Uncertainty and forecasting of water quality* (pp. 259–291). Berlin: Springer.

Fitz, H. C., Voinov, A., & Costanza, R. (1995). The everglades landscape model: Multiscale sensitivity analysis. South Florida Water Management District, Everglades Systems Research Division, West Palm Beach, FL, 88 p.

Fontaine, T. A., & Jacomino, V. M. F. (1997). Sensitivity analysis of simulated contaminated sediment transport. *Journal of the American Water Resources Association, 33*(2), 313–326.

Frey, H. C., & Patil, S. R. (2002). Identification and review of sensitivity analysis methods. *Risk Analysis, 22*(3), 553–578.

Gardner, R. H., Rojder, B., & 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). Nykoping, Sweden.

Gelman, A., Carlin, J., Stern, H., & 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., & Sorooshian, S. (1985a). The automatic calibration of conceptual watershed models using derivative-based optimization algorithms. *Water Resources Research, 21*(4), 473–485.

Gupta, V. K., & Sorooshian, S. (1985b). The relationship between data and the precision of parameter estimates of hydrologic models. *Journal of Hydrology, 81*, 57–77.

Haimes, Y. Y. (1998). *Risk modelling, assessment and management*. New York: Wiley.

Harlin, J., & Kung, C.-S. (1992). Parameter uncertainty in simulation of design of floods in sweden. *Journal of Hydrology, 137*, 2009–2230.

Hendrickson, J. D., Sorooshian, S., & Brazil, L. E. (1988). Comparison of Newton-type and direct-search algorithms for calibration of conceptual rainfall-runoff models. *Water Research Research, 24*(5), 691–700.

Holling, C. S. (1978). *Adaptive environmental assessment and management*. Chichester, UK: Wiley.

Ibrekk, H., & Morgan, M. G. (1987). Graphical communication of uncertain quantities to nontechnical people. *Risk Analysis, 7*(4), 519–529.

Isaaks, E. H., & Srivastava, R. M. (1989). *An introduction to applied geostatistics*. New York: Oxford University Press.

Jaffe, P. R., Paniconi, C., & Wood, E. F. (1988). Model calibration based on random environmental fluctuations. *Journal of Environmental Engineering, 114*(5), 1136–1145.

Jain, S., & Lall, U. (2001, December). Floods in a changing climate: Does the past represent the future? *Water Resources Research, 37*(12), 3193–3205.

Jensen, F. V. (2001). *Bayesian networks and decision graphs*. New York: Springer.

Kann, A., & Weyant, J. P. (1999). A comparison of approaches for performing uncertainty analysis in integrated assessment models. *Journal of Environmental Management and Assessment, 5*(1), 29–46.

Kelly, E., & Campbell, K. (2000, February). Separating variability and uncertainty-making choices. *Human and Ecological Risk Assessment, An International Journal* (in press).

Kelly, E. J., Campbell, K., & Henrion, M. (1997, November). 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.

Kelly, E., & Roy-Harrison, W. (1998). A mathematical construct for ecological risk: A useful framework for assessments. *Human and Ecological Risk Assessment, An International Journal, 4*(2), 229–241.

Lal, A. M. W. (2000, May). Numerical errors in groundwater and overland flow models. *Water Resources Research, 36*(5), 1237–1248.

Lal, W., Obeysekera, J., & van Zee, R. (1997, August). Sensitivity and uncertainty analysis of a regional simulation model for the natural system in South Florida, in managing water: Coping with scarcity and abundance. In *Proceedings, 27th Congress of International Association for Hydraulic Research*. San Francisco, CA.

Lemons, J. (Ed.). (1996). *Scientific uncertainty and environmental problem solving*. Cambridge, MA: Blackwell.

Lopez, A., & Loucks, D. P. (1999). *Uncertainty representations in water quality prediction. Manuscript, civil and environmental engineering*. Ithaca, NY: Cornell University.

Loucks, D. P., Stedinger, J. R., & Haith, D. A. (1981). *Water resource systems planning and analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Ludwig, D., Hilborn, R., & Walters, C. (1993). Uncertainty, resource exploitation, and conservation: Lessons from history. *Science, 260*, 17–36.

Majoni, H., & Quade, E. S. (1980). *Pitfalls of analysis*. New York: Wiley.

McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J., 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, Published for the Intergovernmental Panel on Climate Change. http://www.ipcc.ch/pub/tar/wg2/index.htm

McCuen, R. H., & Snyder, W. M. (1983). *Hydrologic modeling: Statistical methods and applications*. Prentice-Hall, NJ: Englewood Cliffs.

Meixner, T., Gupta, H. V., Bastidas, L. A., & Bales, R. C. (1999). Sensitivity analysis using mass flux and concentration. *Hydrological Processes, 13*, 2233–2244.

Miser, H. J. (1980). Operations research and systems analysis. *Science, 209*, 174–182.

Morgan, M. G., & Henrion, M. (1990). *Uncertainty, A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge, UK: Cambridge University Press.

National Research Council (NRC). (1996). Committee on risk characterization, understanding risk: Informing decision in a democratic society (P. S. Stern & H. V. Fineberg, Eds.). Washington, D.C.: National Academy Press.

Nearing, M. A., Deer-Ascough, L., & Laflen, J. F. (1990). Sensitivity analysis of the WEPP hillslope profile

erosion model. *Transactions of the American Society of Agricultural Engineers, 33*(3), 839–849.

Phillips, D. L., & Marks, D. G. (1996). Spatial uncertainty analysis: Propagation of interpolation errors in spatially distributed models. *Ecological Modelling, 91*, 213–229.

Press, S. J., & Tanur, J. M. (2001). *The subjectivity of scientists and the Bayesian approach*. New York: Wiley.

Reckhow, K. H. (1994). Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. *Ecological Modelling, 72*, 1–20.

Reckhow, K. H. (1999). Water quality prediction and probability network models. *Canadian Journal of Fisheries and Aquatic Sciences, 56*, 1150–1158.

Reckhow, K. (2002, January). *Applications of water models: Prediction uncertainty and decision making*. Presentation at Model Uncertainty Workshop, South Florida Water Management District, West Palm Beach, FL.

Saltelli, A., Chan, K., & Scott, E. M. (Eds.). (2000). *Sensitivity analysis*. Chichester, UK: Wiley.

Schweppe, F. C. (1973). *Uncertain dynamic systems*. Englewood Cliffs, NJ: Prentice-Hall.

Shapiro, H. T. (1990). The willingness to risk failure. *Science, 250*(4981), 609.

Simon, I.-I. A. (1988). *Prediction and prescription in system modeling, 15th anniversary of IIASA*. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Sklar, F. H., & Hunsaker, C. T. (2001). The use and uncertainties of spatial data for landscape models: An overview with examples from the Florida Everglades, Chapter 2 in spatial uncertainty in ecology, implications for remote sensing and GIS applications (Hunsaker et al., Eds.). Berlin: Springer.

Sorooshian, S., Duan, Q., & Gupta, V. K. (1993). Calibration of rainfall-runoff models: Application of global optimization to the Sacramento soil moisture accounting model. *Water Resources Research, 29*(4), 1185–1194.

Sorooshian, S., Gupta, V. K., & 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, 19* (1), 251–259.

Soutter, M., & Musy, A. (1999). Global sensitivity analyses of three pesticide leaching models using a Monte-Carlo approach. *Journal of Environmental Quality, 28*, 1290–1297.

Spear, R. C., & Hornberger, G. M. (1980). Eutrophication in peel inlet—II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research, 14*, 43–49.

Stokey, E., & Zeckhauser, R. (1977). *A primer for policy analysis*. New York, NY: W. W. Norton and Co., Inc.

Suter, G. W., II. (1993). *Ecological risk assessment*. Boca Raton, FL: Lewis Publishers.

Tattari, S., & Barlund, I. (2001). The concept of sensitivity in sediment yield modelling. *Physics and Chemistry of the Earth, 26*(1), 27–31.

van Griensven, A., Francos, A., & Bauwens, W. (2001). Sensitivity analysis and auto-calibration of an integral dynamic model for river water quality (in progress).

van Harn Adams, B. A. (1998). *Parameter distributions for uncertainty propagation in water quality modeling* (Doctoral dissertation). Duke University.

van Straten, G. (1983). Maximum likelihood estimation of parameters and uncertainty in phytoplankton models. In M. B. Beck & G. van Straten (Eds.), *Uncertainty and forecasting of water quality*. New York: Springer.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge, UK: Cambridge University Press.

Warwick, J. J., & Cale, W. G. (1987). Determining likelihood of obtaining a reliable model. *Journal of Environmental Engineering, 113*(5), 1102–1119.

Exercises

8.1 Distinguish between sensitivity analysis and uncertainty analysis.

8.2 Consider the allocation model used in previous chapters involving three water consumers i. Allocations x_i of water can be made from a given total amount Q to the three consumers. The respective benefits are $(6x_1 - x_1^2)$, $(7x_2 - 1.5x_2^2)$ and $(8x_3 - 0.5x_3^2)$. Discuss possible sources of uncertainty in model structure and model output, and identify and display parameter sensitivity.

8.3 Discuss how model output uncertainty is impacted by both model input uncertainty as well as parameter sensitivity.

8.4 In many water resources studies considerable attention is given to the uncertainty of water supplies (precipitation, streamflows, evaporation, infiltration, etc.) and much less attention is given to the uncertainty of the management objectives, the costs and benefits of infrastructure, the political support associated with alternative possible decisions, and the like. Develop a simple water resources planning model involving the management of water quantity and quality and show how these management objective uncertainties may actually dominate the hydrologic uncertainties.

8.5 Perform a deterministic sensitivity analysis for the consumer 1 in Exercise 8.2. Consider the three parameters, Q, 6 and 1; the latter two numbers are the parameters of the benefit function. Low values of these three parameters are 3, 3, and 0.5, respectively. Most likely values are 6, 6, and 1. High values are 12, 9, and 1.5. Display the results using a Pareto chart, a tornado diagram, and a spider plot.

8.6 Referring to water allocation problem defined in Exercise 9.2, assume the available amount of water Q is uncertain. Its cumulative probability distribution is defined by $q/(6 + q)$ for values $q \geq 0$ of the random variable Q. The expected value of Q does not exist. Perform an uncertainty analysis showing how to define, at least approximately:

- Estimating the mean and standard deviation of the outputs.
- Estimating the probability the performance measure will exceed a specific threshold.
- Assigning a reliability level on a function of the outputs, e.g., the range of function values that is likely to occur with some probability.
- Describing the likelihood of different potential outputs of the system.

Show the application of Monte Carlo sampling and analysis, Latin hypercube sampling, generalized likelihood uncertainty estimation and factorial design methods.

Water resource systems typically provide a variety of economic, environmental, and ecological services. They also serve a variety of purposes (e.g., water supply, flood protection, hydropower production, navigation, recreation, and waste assimilation 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 alternative system performances. Some of these performance criteria may be conflicting. In these cases tradeoffs exist among conflicting criteria and these tradeoffs should be considered when searching for the best compromise. This chapter presents ways of identifying and working with these performance criteria in the political process of selecting the best decision.

9.1 Introduction

Decision-makers and those who influence them are people, and people's opinions and experiences and goals may differ. These differences force one to think in terms of tradeoffs. Decisions in water resources management inevitably involve making tradeoffs among competing opportunities, goals or objectives. One of the tasks of water resource 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 the largely political process involving all interested stakeholders to find the best compromise decision.

If every system performance measure or objective could be expressed in the same units, and if there was only one decision-maker and one objective or goal, then decision-making would be relatively straight forward. Such is not the case when dealing with the public's water resources.

A 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 net benefits, i.e., the benefits less the costs. If the maximum difference between benefits and costs was positive, that was the best plan or policy. But not all system performance criteria we consider today 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 involving 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 likely agree on to how best to allocate or share those net benefits among them.

Clearly, water resource planning and management takes place in a multiple criteria environment. A key element of many problems facing designers and managers is the need to deal explicitly with multiple ecological, economical, and social impacts, expressed in multiple metrics that may result from the design, management,

and operation of water resource systems. 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 (Hipel et al. 2015).

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 only complicates the life of professional planners and managers. Increasingly important sources of information come from discussion groups, public hearings, negotiations, and dispute-resolution processes. Using the types of modeling methods discussed in this chapter can potentially inform the debates that occur in these meetings. In addition, the application of game theory may be helpful in reaching a consensus among stakeholders having different objectives and desires (Madani 2010; Madani and Hipel 2011).

Eventually someone or some organization must make a decision. Even if water resource analysts view their job as one of providing options or tradeoffs for someone else to consider, even they are making decisions that define or limit the range of those options or tradeoffs. The importance of making informed effective decisions applies to everyone.

9.2   Informed Decision-Making

Informed decision-making involves both, qualitative thinking and analysis as well as quantitative modeling (Hammond et al. 1999). Qualitative thinking and analysis typically follows quantitative analyses. Qualitative analyses are 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 socially acceptable alternatives that address and meet each objective, and
- the key tradeoffs among all interests and objectives involved.

When the objectives and general alternative ways of meeting these objectives are not well defined, no amount of quantitative modeling and analysis will make up for this weakness. The identification of objectives and general alternatives are the foundation upon which water resource 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 that no one has thought of before. 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 might be reduced by increasing surface water reservoir storage capacity or by increasing 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 the cost of increased reservoir storage and pumping capacities and the increased reliability of meeting the supply target. However these models will not identify and evaluate completely different alternatives, such as importing water by trucks or the use of canals or pipelines from other river basins, implementing water use restrictions, or water reuse, unless of course those options are included in those models. Someone has to think of these general alternatives before models can help identify just how many trucks or the capacities of canals or pipes or how to best implement water reuse and for what uses.

Time needs to be taken to identify the relevant objectives and general alternatives that then can be included in quantitative analyses. Clearly some preliminary screening of general alternatives is and should be carried out at the

qualitative level, but the alternatives that remain (assuming the best is not obvious) should then be further analyzed using quantitative modeling methods. This includes the methods of quantifying qualitative objectives and constraints discussed in Chap. 5.

Being creative in the identification of 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. For example, 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 a target water allocation. Economic, environmental, ecological, and purely physical objectives such as these are able to 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 modeling is employed to identify more precisely the design and operation of structural and nonstructural alternatives that best satisfy system performance criteria, and the impacts and tradeoffs among these 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 not? 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?

Thus the modeling process ends with some more qualitative study. Models do not replace human judgment. Humans, not quantitative models, are responsible for water resources planning and management decisions as well as the decision-making process itself that identifies performance criteria and general alternatives.

9.3 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 wellbeing. 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.

By asking "how" any specified broad fundamental criterion or objective can be achieved usually leads, eventually, to more specific system performance criteria or objectives and to the means of improving these criteria, i.e., to the general system design and operating policy alternatives themselves. As one gets further from the fundamental objective that most will agree to, there is a greater chance of stakeholder disagreement. For example we might answer the question "How can we maximize 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 downstream flow augmentation from upstream reservoirs 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 nonpoint 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 may not have been 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, it 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, it is a possible sign there are other objectives and alternatives waiting to be identified and evaluated. What are they? Get them identified. Consider them along with the others that have been already defined.

More fundamental objectives can be identified by asking the question "why?" Answers to the question "why?" 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.

9.3.1 Constraints on Decisions

Constraints limit alternatives. There are some laws of physics that obviously we cannot change. Water will naturally flow downhill unless of course we pump it uphill. Society can limit what can be done to satisfy any performance criterion as well, but it is not the time to worry about them during the process of objective and general alternative identification. Be creative, and don not get stuck in the status quo, i.e., 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 alternatives, but that can be determined later. When performing qualitative exercises to identify objectives and general alternatives enlarge the number of options and think creatively. If something seems really worth further consideration, then 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—again 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 Great Lakes, the Sacramento-San Joaquin Delta, and Florida Everglades regions in the US.

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 when identifying scenarios. Consider the whole spectrum of possible events that your decision responds to. 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 also plan for.

It is 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 actions.

Finally, target values of objectives and goals should not be set too low (easy to meet). The chances of finding good unconventional alternatives are 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.

9.3.2 Tradeoffs Among Performance Criteria

Tradeoffs are inevitable when there are conflicting 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, lakes, rivers and streams, 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, that compete for those same, and often limited, amounts of money.

Tradeoffs arise because we all want more of good things, and many of these good things are conflicting. We cannot have cleaner water in our homes or in our environment without spending money, and minimizing costs is always worthy of consideration. Identifying these tradeoffs is one of the tasks of qualitative as well as quantitative analyses. Qualitatively we can identify what the tradeoffs will involve, e.g., cleaner water in our streams and rivers require increased costs. Just how much money 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 nondominated tradeoffs is obviously helpful. 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 that can identify these tradeoffs among quantitative objectives cannot go the next step, i.e., identify the best compromise decision. While models can help identify and evaluate alternatives, they cannot take the place of human judgments that are needed to make the final decision—the selection of the best tradeoffs.

Models can help identify nondominated tradeoffs among competing objectives or system performance indicators. These are sometimes called efficient tradeoffs. Efficient decisions are those that cannot be altered so as to gain more of one objective without worsening one or more other 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 considerations, since why would any rational individual prefer such decisions when there are better ones available with respect to all the objectives being considered. However, some may consider a quantitatively nondominated or efficient solution inferior and dominated with respect to one or more other objectives that were omitted from the analysis. Eliminating inferior or dominated alternatives from the political decision-making process may not always be very helpful. One of these inferior alternatives could be viewed as the best by some who are considering other criteria either not included or, in their opinions, not properly quantified in such analyses.

Tradeoffs exist not only among conflicting outcome objectives, system performance

indicators, or impacts. Tradeoffs can also exist 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 modeling 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.

9.4   Quantifying Performance Criteria

So far this chapter has focused on the critical qualitative aspects of identifying objectives and general alternatives. The remainder of this chapter will focus on quantifying various criteria and how to use them to compare various alternatives and to 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 mine 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 watershed conservation program could have a flood control objective quantified as the height of the protected flood stage. It could have a regional development objective quantified as increased income. Quantification does not require that monetary costs and benefits be assigned to all objectives.
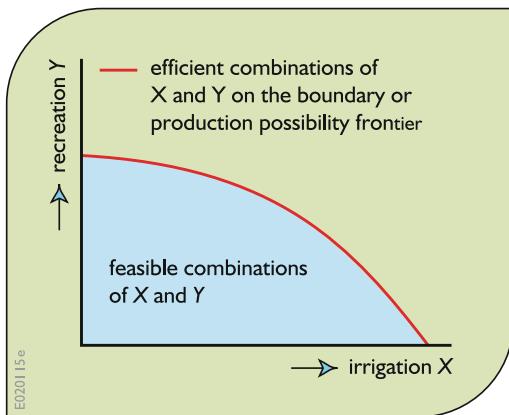
The following subsections review various economic, environmental, ecological, and social criteria.

9.4.1   Economic Criteria

Water resource system development and management is often motivated by economic criteria. It goes without saying that money is important; it's completing uses often makes it a limiting resource Most reservoirs, canals, hydropower facilities, groundwater-pumping systems, locks, and flood control structures have been built and are operated for economic reasons. The benefits and costs of such infrastructure can be expressed in monetary units. Typical objectives 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.

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. Irrigation and recreation are not very compatible. Recreation benefits are greater when reservoir elevations remain high throughout the summer recreation season, just when satisfying an irrigation demand that exceeds the inflow would normally cause a drop in the reservoir storage level. Thus the project has two conflicting purposes: provision of irrigation water and of recreation opportunities.

Fig. 9.1 A two-purpose planning problem showing the feasible and efficient combinations of irrigation water X and recreation visitor days Y
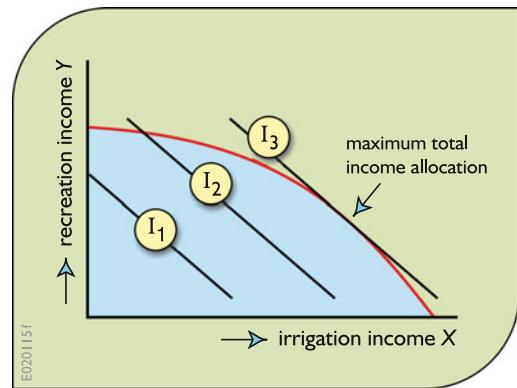
Fig. 9.2 Two-purpose project involving tradeoffs between irrigation and recreation income and various constant total income levels, $I_1 < I_2 < I_3$

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 Fig. 9.1. The solid line in Fig. 9.1, 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 (i.e., by 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 on the frontier. They are said to be *technologically efficient* in the sense that more of either X or Y cannot be obtained without a 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 satisfy various demands for both X and Y.

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

$$I = p_X X + p_Y Y \qquad (9.1)$$

Values of X and Y that result in fixed income levels $I_1 < I_2 < I_3$ are plotted in Fig. 9.2. The value of X and Y that maximizes the entrepreneur's income is indicated by the point 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 both should, under competitive conditions, produce the same result.

When applied to water resources planning, benefit-cost analysis presumes 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.

9.4.1.1 Benefit and Cost Estimation

In a 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, i.e., 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, labor 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 nonexistent, but for which 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 is easily 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 $X_1, X_2, \ldots,$ 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 $D(X_j)$,

$$W(X_j) = B(X_j) - D(X_j) \qquad (9.2)$$

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 this compensation to be paid. Those who lose favorite scenic sites or the opportunity to use a wild river or who must hear the noise or breathe dirtier air or who suffer a loss in their property value 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, i.e., 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 these distributional issues.

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 individual it is very difficult not to provide them to others. Collective goods often also have the property that their non-consumptive use by one person does not prohibit or infringe upon their use by others.

Community flood protection is an example of a public good. Once protection is provided for one individual, it is often 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 (such as the taxpayers), 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 due to the project. The appropriate method for benefit and cost estimation depends on whether or not the market prices reflecting true social values are available or if such prices can be constructed.
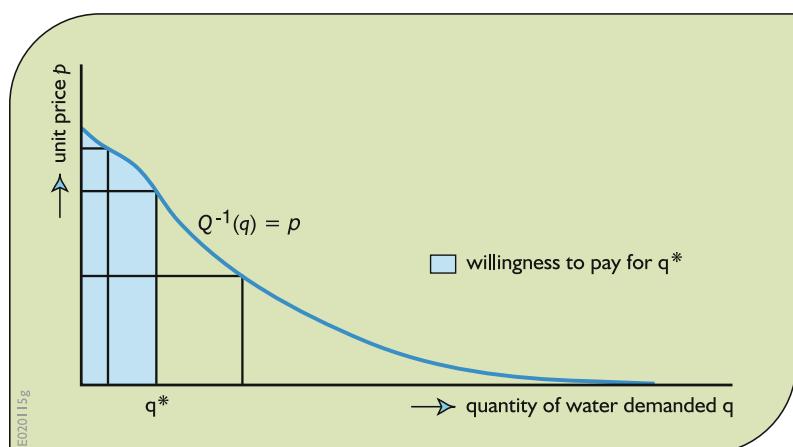
Market Prices Equal Social Values

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 Fig. 9.3, the lower the unit price, the more water individuals are willing to buy, i.e., the greater will be the demand.

The willingness to pay a given unit price p is defined by the area under the demand curve. As Fig. 9.3 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

Fig. 9.3 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^*

of water q^* is the area under the demand curve from $Q = 0$ to $Q = q^*$.

$$\text{Willingness to pay for } q^* = \int_0^{q*} Q^{-1}(q)dq$$
(9.3)
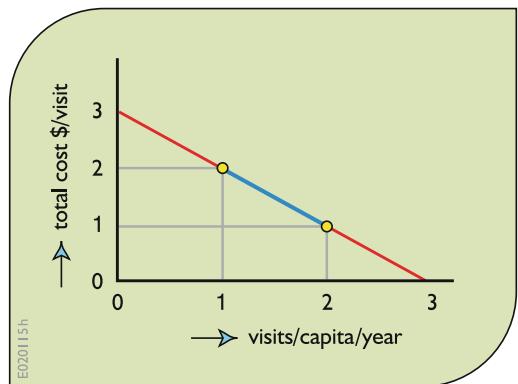
Consumer's willingness to pay for a product is an important concept in welfare economics.

Market Prices not Equal to Social Values

It frequently occurs that market prices do not truly reflect the true social value of the various 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 or 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 in the absence of the project's adoption, 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 centers. Center A has a population of 10,000 and the more distant Center B has a population or 30,000. From questionnaires it is estimated that if
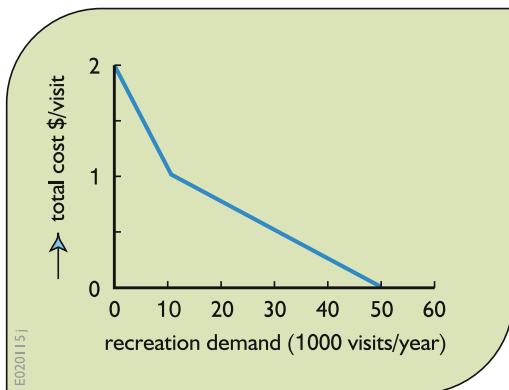
Fig. 9.4 Estimated relationship between travel cost and visits per capita

access to the recreation area is free, 20,000 visits per year will be made from Center A at an average round-trip travel cost of $1. Similarly 30,000 visits per year will come from Center 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 Fig. 9.4, 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 center A) at a cost of $1/visit, and an average of one visit per capita (from Center B) at a cost of $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, it is possible to 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 Center A would then be $2 per visit. From Fig. 9.4 at $2 per visit, one visit per capita is made; hence 10,000 visits per year can be expected from Center A. The resulting cost to users from Center B is $3 per visit, and hence from Fig. 9.4 no visits would be expected. Therefore, one point on the demand curve (10,000 visits at $1) is obtained. Similarly,

Fig. 9.5 Demand function for recreation facility

it can be inferred that at a $2 admission price there will be no visits from either center. 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 Fig. 9.5. 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 illustrative only. The average cost of travel time (which differs for each population center) 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 esthetics 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. Their shadow prices or dual variables (indicating the marginal cost associated with a unit change in the regulation or requirement) is likely to be as close to monetizing such non-monetary impacts as one can get, yet recognizing the actual monetary benefits could be greater. 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 esthetic) objectives, if quantified, can be incorporated into a multiobjective decision-making process. However, this falls short of the assignment of monetary benefits that is often possible for the first three categories of benefits.

9.4.1.2 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 unlikely reflects 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 labor is the marginal value of leisure forgone, since there is no alternative productive use of those individuals. Yet, the labors 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 multiobjective nature of the decisions. One must

consider environmental impacts, income redistribution effects, and a host of other local, , and national goals.

Other important considerations are the financial, technical, and political feasibilities of alternative plans. Particularly important when plan are proposed by government agencies is the relative political and legal clout of those who support the plans and those who oppose them. Still, a plan's economic efficiency is an important measure of a plan's value to society and often serves as an indicator of whether it should be considered at all.

9.4.1.3 Long- and Short-Run Benefit Functions

When planning the capacities and target values associated with water resource development projects, it is often convenient to think of two types of benefit functions: the long-run benefits and the short-run benefits. In *long-run planning,* the capacities of proposed facilities and the target allocations of flows to alternative uses or target storage volumes in reservoirs 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 over time. In *short-run planning,* the capacity of facilities and any associated targets are assumed known. The problem is one of managing or operating a given or proposed system under varying supply and demand conditions.

For example, if a water-using firm is interested in building a factory requiring water from 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 the capacity of that factory, the number of employees hired, and the amount of product produced, etc. 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. Whatever the target, if it is satisfied, long-run benefits result, Short-run benefits are the benefits one actually
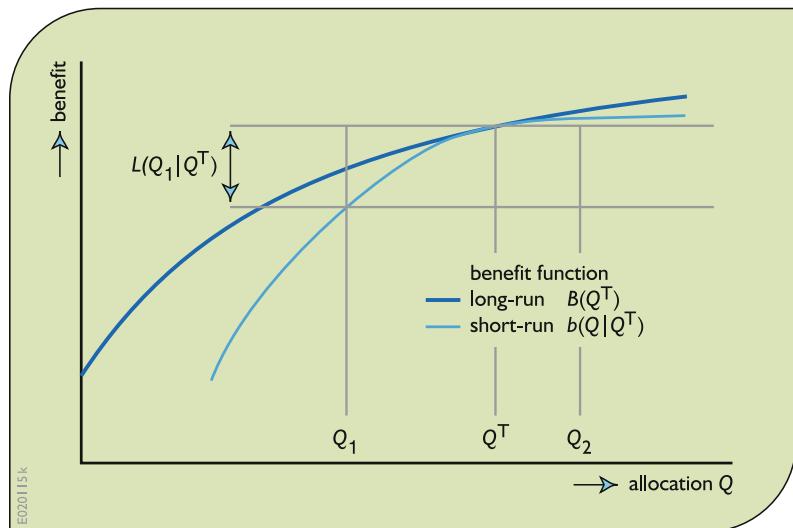
can obtain by operating a system having fixed capacities and target values. If the water resources available in the short run are those that can meet all the targets that were established when long-run decisions were made, then estimated long-run benefits can be achieved. Otherwise the benefits actually 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 and demand conditions.

The distinction between long-run and short-run benefits can be illustrated by considering again a potential water user at a particular site. 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 $B(T)$ and the short-run net benefits $b(Q|Q^T)$ for a particular target Q^T is illustrated in Fig. 9.6.

The long-run function $B(Q^T)$ in Fig. 9.6 reflects the benefits 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 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 Fig. 9.6. In other words the short-run

Fig. 9.6 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

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 Fig. 9.7. 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.

$$L(Q|Q^T) = B(Q^T) - b(Q|Q^T) \qquad (9.4)$$

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
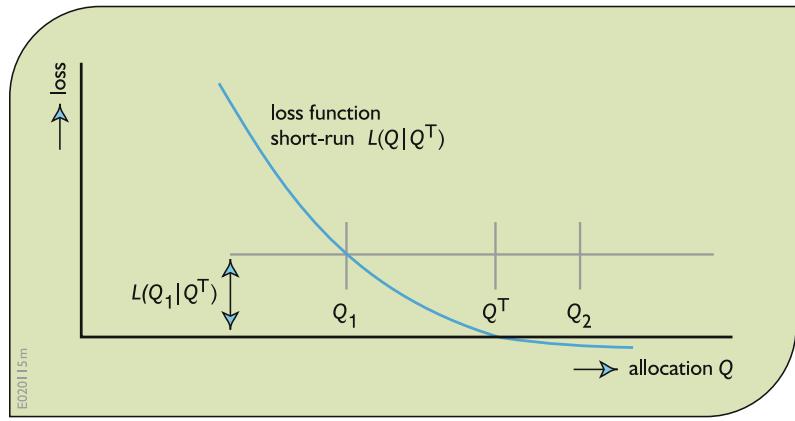
$$Q = Q^T - D + E \qquad (9.5)$$

Denote the loss function for a deficit allocation as $L^D(D)$ and for a surplus allocation, $L^E(E)$. As indicated above, the later may be a negative loss, i.e., a gain, at least for some range of E, as shown in Fig. 9.7.

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 it 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

Fig. 9.7 Short-run loss function, $L(Q|Q^T)$, associated with an allocation Q and target allocation Q^T

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 $C_j(K_j)$.

$$NB = \sum_j \left[B_j(T_j) - L_j(Q_j|T_j) - C_j(K_j) \right] \quad (9.6)$$

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, etc., some of which cannot easily be transferred to other uses. At least on an annual basis, most farmers reappraise these resource allocations in light of the projected market prices of the generated commodities and the availability and cost of water, energy, labor 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 land, water, capital and labor. 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 modeling, however, this planning hierarchy can generally be described by two levels, denoted as long-run 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 modeled. 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 independent of any other period's allocation, at least very easily. The benefits from irrigation come only when the crops are harvested, e.g., at

the end of the last period of each growing season. In this case some mechanism is needed to determine the benefits obtained from a series of allocations over time, as will be presented in the next chapter.

9.4.2   Environmental Criteria

Environmental criteria for water resource projects can include water quantity and quality conditions. These conditions 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 esthetic 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 and dissolved oxygen would normally have a minimum acceptable concentration. These limiting concentrations and their specified reliabilities are often based on standards established by governmental or international environmental or health 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.

Few 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." Stakeholder membership functions can define what is considered virtually free, and these can be included as objectives or constraints in models (as discussed in Chap. 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 resource planners and managers can also relate to recreational and land use activities. These typically address hydrologic 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 boaters 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, say 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.

9.4.3   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 impact 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. (Like people, it's hard to satisfy everyone all the time.) If one species feeds on another, it is hard to imagine how to maximize the health of both. The conditions that favor one species group may not favor another. Hence variation in habitat conditions is important for the sustainability of both, and indeed for achieving resilient biodiverse ecosystems.

While ecosystem habitats exhibit more diversity when hydrologic conditions vary, as in

nature, than when they are relatively constant, hydrologic variation is often not desired by many human users of water resource systems. Reducing hydrologic variation and increasing the reliability of water resource systems has often been the motivating factor in the design and operation of hydraulic engineering works.

The state of ecological habitats is in part functions of how water is managed. One way to develop performance indicators of ecosystems is to model the individual species making up the ecosystem, or at least a subset of important indicator species. This is often difficult. Alternatively one can define habitat suitability indices for these important indicator species. This requires (1) selecting the indicator species representative of each particular ecosystem, (2) identifying the hydrological attributes that affect the wellbeing of those indicator species during various stages of their life cycles, and (3) quantifying the functional relationships between the wellbeing 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 unfavorable.

Examples of hydrological attributes that impact 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 hydrologic 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.
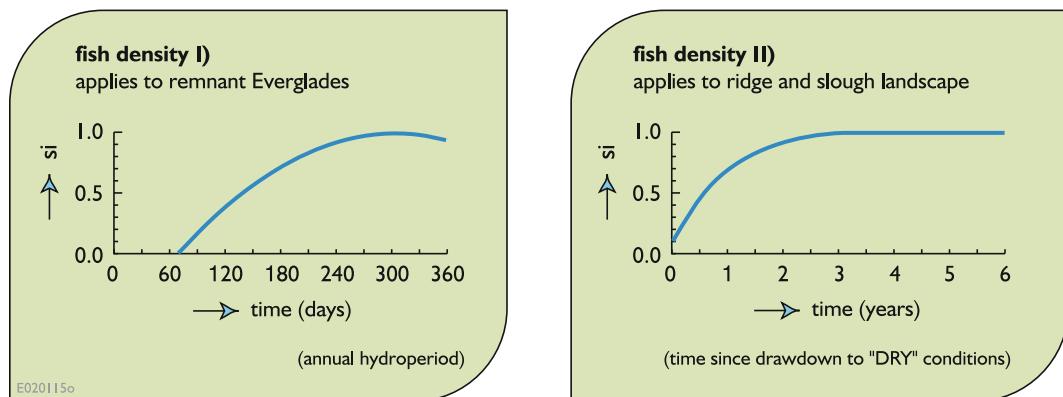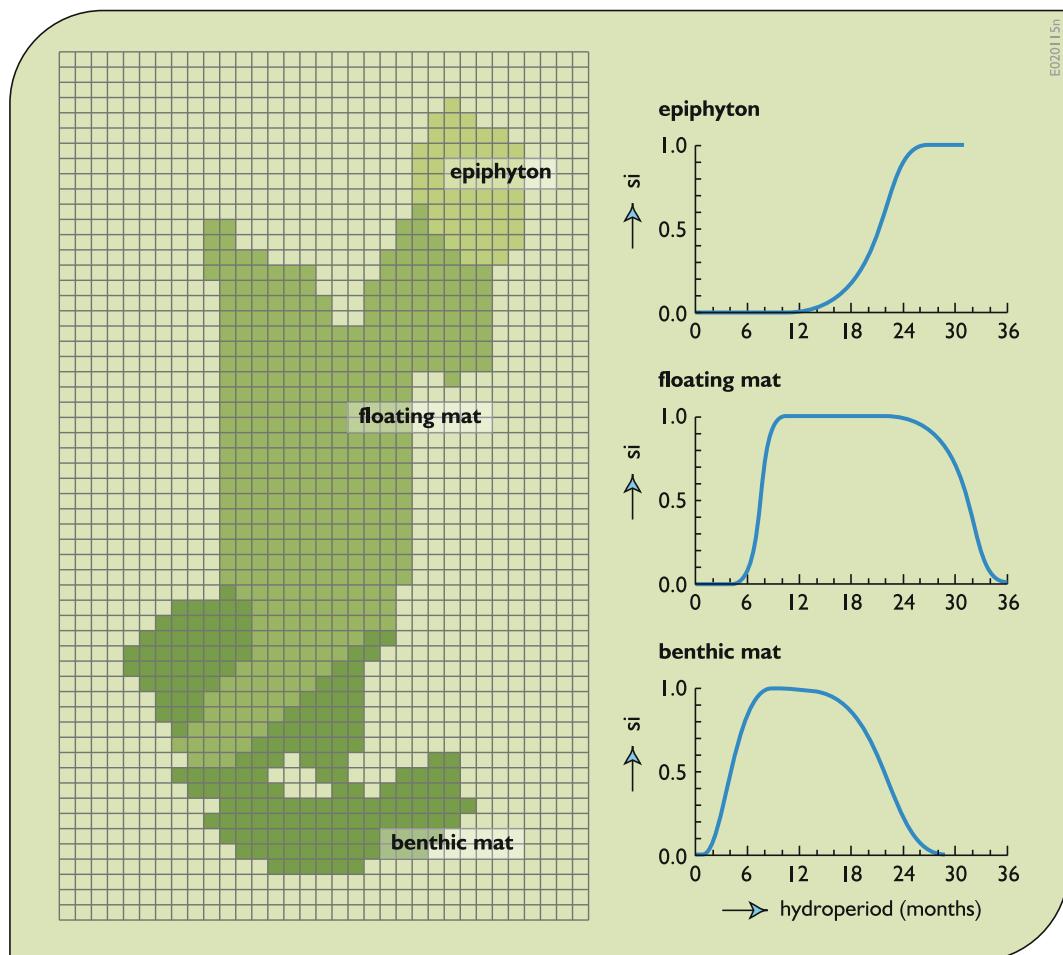
Figure 9.8 illustrates three proposed habitat suitability indices for periphyton (algae) and fish in parts of the Everglades region of southern Florida in the US. All are functions of

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.
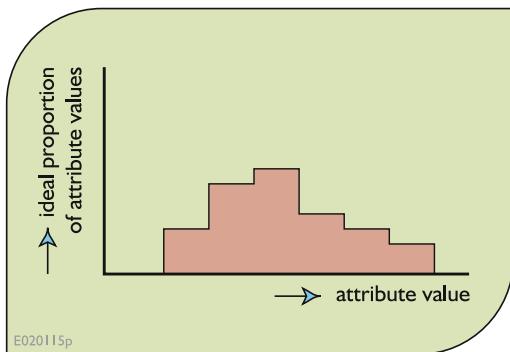
There are other functions that would influence the growth of periphyton and fish, such as the concentrations of phosphorus or other nutrients in the water. These are not shown. Figure 9.8 merely illustrates the construction of habitat suitability indices.

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 one 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 hydrologic 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 if so the applicable segments are understood to cover those ranges.) Such a function is shown in Fig. 9.10. Figure 9.9 happens to be a discrete distribution, but it could have been a continuous one as well.
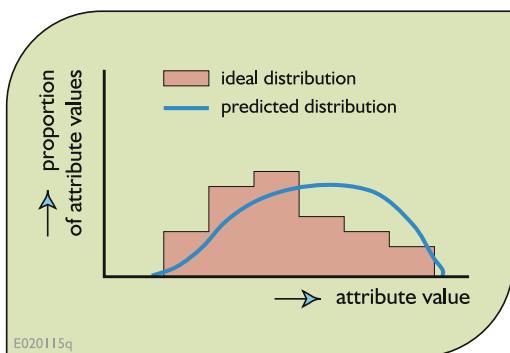
Any predicted distribution of attribute values resulting from a simulation of a water management policy can be compared to this ideal distribution, as is shown in Fig. 9.10. The fraction of overlapping areas of both distributions is an indication of the suitability of that habitat for that indicator species.

Fig. 9.8 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 US

Fig. 9.9 Ideal range of values of a hydrologic attribute for a particular component of an ecosystem

Fig. 9.10 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

$$
\begin{aligned}
\text{Habitat suitability} = &\text{Fraction of area under}\\
&\text{the ideal and simulated}\\
&\text{distributions of attribute values.}
\end{aligned}
$$
(9.7)

In Fig. 9.10, this is the red shaded area under the blue curve.

To identify a representative set of indicator species of any ecosystem, the hydrologic attributes or 'stressors' that impact those indicator species, and finally the specific functional relationships between the hydrologic 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 nontrival task of trying to simplify it to define habitat suitability performance criteria. However once identified, these habitat suitability performance criteria can give water resource 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.

9.4.4   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 resource system performance. Economic, environmental, and ecological impacts resulting from water management policies directly affect people. 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 labor, for example.

Another social performance criterion is human settlement displacement. The number of families that must move from their homes because of, for example, flood plain restoration or reservoir construction, is always of concern. These impacts can be expressed as a function of the extent of flood plain restoration or reservoir storage capacity, respectively. Often the people most affected are in the lower-income groups, and this raises legitimate issues of social 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 flood plain or wetland? Conversely, if social objectives are first identified, by asking and then answering 'how' they can be achieved 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. In times of social stress due to, for example, floods, droughts, or disease caused by waterborne bacteria, viruses, and pollutants, press coverage and public involvement often 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 disinterested 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 such planning for more stressful conditions should take place.

9.5   Multicriteria 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 impacted stakeholder the happiest. When agreement is not universal and when some objectives conflict with others, we can identify the efficient tradeoffs among the objective values each stakeholder would like to have. In this section some ways of identifying efficient tradeoffs are reviewed. These methods of multicriteria or multiobjective analyses are not designed to identify the best solution, but only to provide

information on the tradeoffs among conflicting quantitative performance criteria. Again, any final decision will be based on qualitative as well as this quantitative information in a political and social process, not by or in a computer.

Even if the same units of measure, e.g., monetary ones, can be used for each performance measure or objective, it is not always appropriate to simply sum them together into a single measure or objective that can be maximized or minimized. Consider for example a water resources development project to be designed to maximize net economic benefits. In the US this is sometimes designated the national economic development (NED) objective. Another objective may be to distribute the costs and benefits of the project in an equitable way. Both objectives are measured in the same 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 will likely agree as to how that pie should be divided up among all the stakeholders. It also matters who pays and who benefits, and by how much. Again, issues of equity and social justice involve judgments, and the challenge of water resource planners and managers, and elected politicians, is to make good judgments. The result is often a plan or policy 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 remaining effluent, regardless of the quality or assimilative capacity of the receiving bodies of water, is one example of this compromise between economic, environmental, and social criteria.

The irrigation-recreation example presented earlier in this chapter illustrates some basic concepts in multiobjective planning. As indicated in Fig. 9.1, one of the functions of multiobjective 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. One needs some way to compare them. Social welfare functions that could provide a basis for comparison is impossible to construct, and the reduction of multiple objectives to a single criterion (as in Fig. 9.2), especially if they are conflicting, 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 *vector optimization* representation of the problem may be applicable. Let the vector X represent the set of unknown decision variables whose values are to be determined and let $Z_j(X)$ 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 objectives $Z_j(X)$ are to be maximized, the model can be written

$$\text{maximize}\left[Z_1(X), Z_2(X), \ldots, Z_j(X), \ldots, Z_J(X)\right]$$

subject to:

$$g_i(X) = b_i \quad i = 1, 2, \ldots, m \qquad (9.8)$$

The objective in Eq. 9.8 is a vector consisting of J separate objectives. The m constraints $g_i(X) = b_i$ define the feasible region of solutions. Again, the vector X represents all the unknown decision variables whose values are to be determined by solving the model.

The vector optimization model is a concise way of representing a multiobjective problem but it is not very useful when trying to solve it. In reality, a vector can be maximized or minimized only if it can be reduced to a scalar. Thus the multiobjective planning problem defined by Model 9.8 cannot, in general, be solved without additional information. Various ways of solving this multiobjective model are discussed in the following subsections.

The goal of multiobjective modeling is the generation of a set of technologically feasible and efficient plans. Recall that an efficient plan is one that is not dominated.

9.5.1 Dominance

A plan 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, X_i, dominates if

$$Z_j(X_i)Z_j(X_k) \quad \text{for all objectives } j \text{ and plans } k \qquad (9.9)$$

and for each plan $k \neq i$ there is at least one objective $j*$ such that

$$Z_{j*}(X_i) > Z_{j*}(X_k) \qquad (9.10)$$

Not very often does one plan dominate all others. If it does, pick it! More often different plans will dominate all plans for different objectives. However, if there exists two plans k and h such that the values of all objectives j for plan k are never less than that for plan h ($Z_j(X_k) \geq Z_j(X_h)$), and for some objective $j*$, plan k provides a higher value than does plan h, ($Z_{j*}(X_k) > Z_{j*}(X_h)$), then plan k dominates plan h and plan X_h can be dropped from further consideration. This assumes of course that all objectives are being considered. If some objectives are not included in the analysis, perhaps because they cannot be quantified, inferior plans with respect to those objectives that are included in the analysis should not be rejected from eventual consideration just based on this quantitative analysis. To work, dominance analysis must consider all objectives. In practice this condition is often impossible to meet.
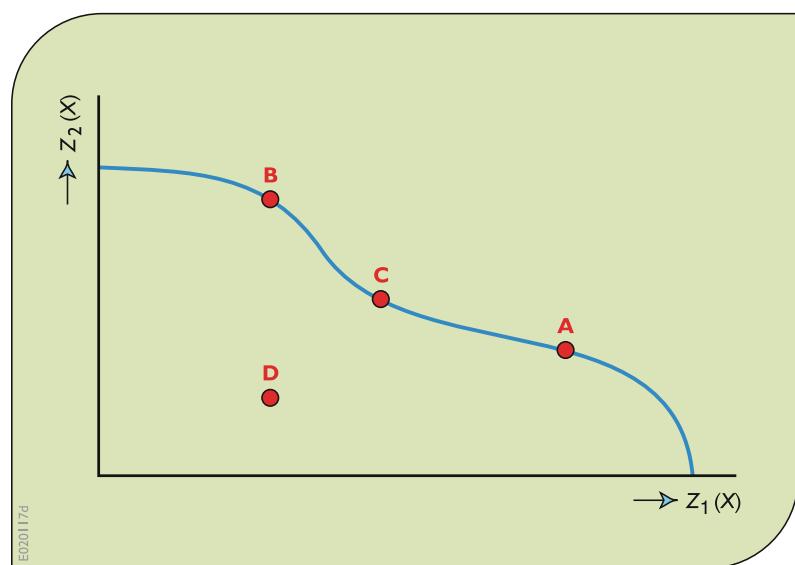
Dominance analysis requires that participants in the planning and management process specify all the objectives that are to be considered. It does not require the assessment of the relative importance of each objective. Non-inferior, efficient, or nondominated solutions are often called

Pareto optimal or Pareto efficient because they satisfy the conditions proposed by the Italian economist and social theorist Vilfredo Pareto (1848–1923). A set of objective values is efficient if in order to improve the value of any single objective, one must accept a diminishment of at least one other objective.

Consider for example three alternatives *A, B,* and *C*. Assume, as shown in Fig. 9.11, that plan *C* is inferior to plan *A* with respect to objective $Z_1(X)$ and also inferior to plan *B* with respect to objective $Z_2(X)$. Plan *C* might still be considered the best with respect to both objectives $Z_1(X)$ and $Z_2(X)$. While plan *C* could have been inferior to both *A* and *B*, as is plan *D* in Fig. 9.11, it should not necessarily be eliminated from consideration just based on a pair-wise comparison. In fact plan *D*, even though inferior with respect to both objectives $Z_1(X)$ and $Z_2(X)$, might be the preferred plan if another objective were included.

Two common approaches for identifying nondominated plans that together identify the efficient tradeoffs among all the objectives $Z_j(X)$ in the Model Eq. 9.8 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 (the blue line in Fig. 9.11).

9.5.2 The Weighting Method

The weighting approach involves assigning a relative weight to each objective to convert the objective vector (in Eq. 9.8) to a scalar. This scalar is the weighted sum of the separate objective functions. The multiobjective Model 9.8 becomes

$$\text{maximize } Z = [w_1Z_1(X) + w_2Z_2(X)\ldots + w_jZ_j(X)\ldots + w_JZ_J(X)]$$
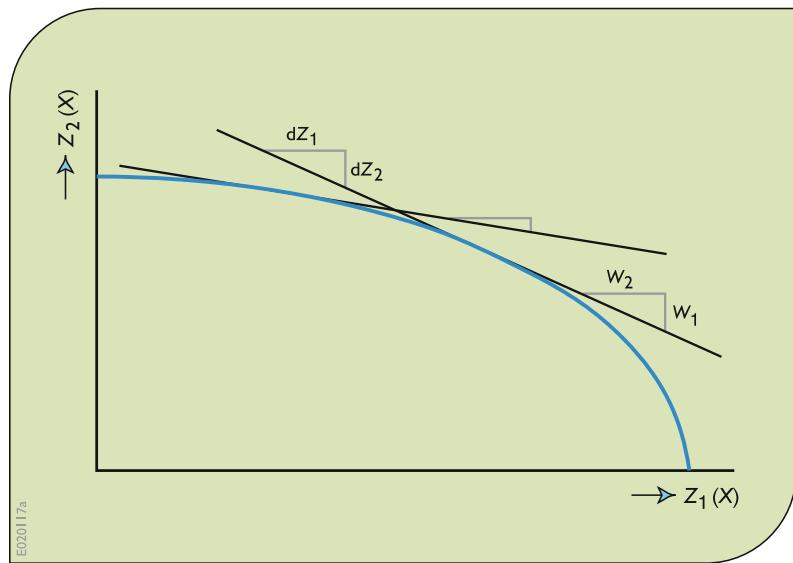
subject to:

$$g_i(X) = b_i \quad i = 1, 2, \ldots, m \qquad (9.11)$$

where the nonnegative 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 functions' 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 X, is

Fig. 9.11 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 nondominated plans. All objectives should be considered before declaring a plan inferior

Fig. 9.12 The efficiency frontier between two objectives, $Z_1(X)$ and $Z_2(X)$, showing the reduction in one objective, say $Z_1(X)$, as the relative weight, w_2, associated with the other objective, increases

$$-\left[\mathrm{d}Z_j/\mathrm{d}Z_k\right]\big|_{x=\text{constant}} = w_k/w_j \qquad (9.12)$$

This applies when each of the objectives is continuously differentiable at the point X in question. This is illustrated for a two-objective maximization problem in Fig. 9.12.

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 judgments. 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 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 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 was made. They will unlikely be the weights that will apply in other places in other circumstances at other times.
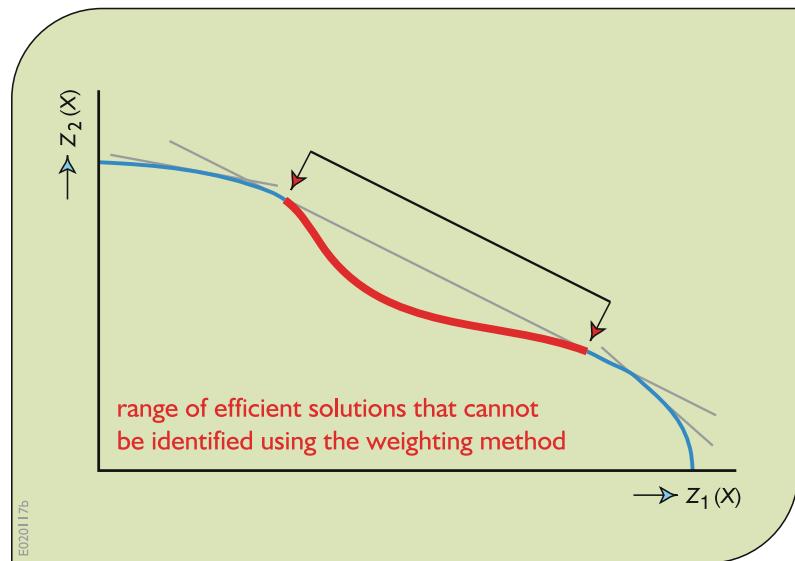
A principal disadvantage of the weighting approach is that it cannot generate the complete set of efficient plans unless the efficiency frontier is strictly concave (decreasing slopes) for maximization, as it is in Figs. 9.1 and 9.12. If the frontier, or any portion of it, is convex, only the endpoints of the convex region will be identified using the weighting method, as illustrated in Fig. 9.13.

9.5.3 The Constraint Method

The constraint method for multiobjective planning can produce the entire set of efficient plans for any shape of efficiency frontier, including that shown in Fig. 9.13, assuming there are tradeoffs among the objectives. In this method one objective, say $Z_k(X)$ 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

Fig. 9.13 An efficiency frontier that cannot be completely identified within its convex region using the weighting method when objectives are being maximized. Similarly for concave regions when objectives are being minimized

range of efficient solutions that cannot be identified using the weighting method

alternative if the lower bounds on the other objective values are binding.

In its general form, the constraint model is

$$\text{maximize } Z_k(X) \qquad (9.13)$$

subject to

$$g_i(X) = b_i \quad i = 1, 2, \ldots, m \qquad (9.14)$$

$$Z_j(X) \geq L_j \quad \forall j \neq k. \qquad (9.15)$$

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* $Z_k(X)$ per unit change in L_j (or $Z_j(X)$ if binding).
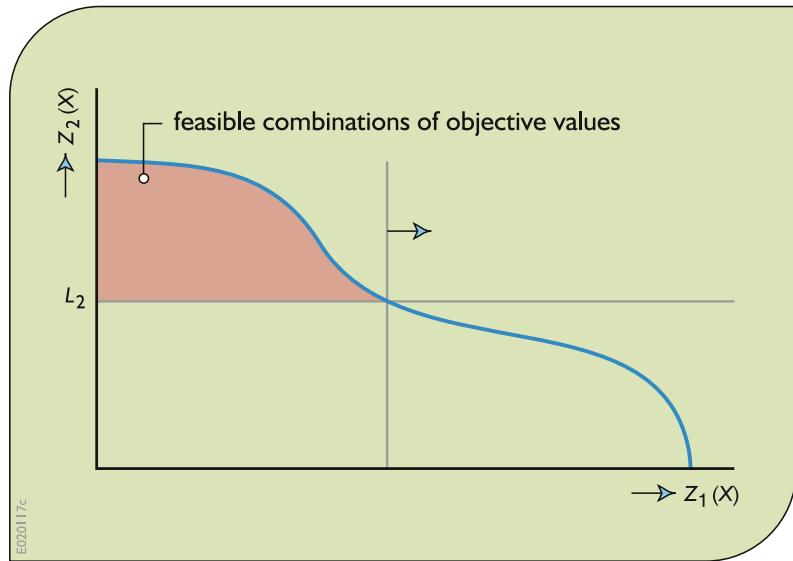
Figure 9.14 illustrates the constraint method for a two-objective problem.

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 nondominated alternatives, as previously demonstrated.

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 stakeholder interest group are not happy in assigning 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

Fig. 9.14 The constraint method for identifying the efficiency frontier by maximizing $Z_1(X)$ while constraining $Z_2(X)$ to be no less than L_2

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 alone will not be sufficient. 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, evaluate and compare each alternative plan. Planners or managers may also need help in identifying which alternatives they prefer. If they are willing to work with analysts, these analysts can help them identify what alternatives they prefer without generating and comparing all the other plans.

There are a number of methods available for assisting in selecting the most desirable non-dominated plan. Some of the more common and simpler ones are described next.

9.5.4 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 that is to be maximized. 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 or a set of plans which the participants in the planning and management process are essentially indifferent. This process is illustrated in Fig. 9.15.
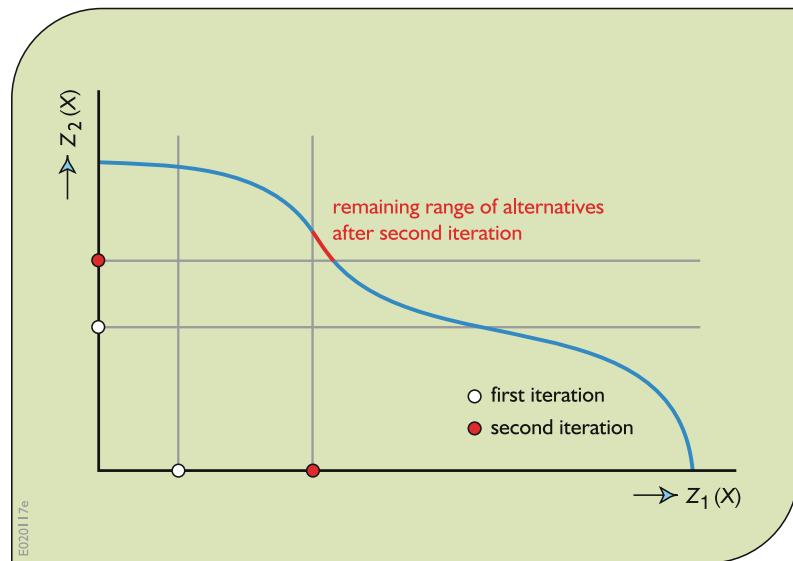
Of course, sometimes the participants in the planning process will be unwilling or unable to sufficiently narrow down the set of available non-inferior plans with the iterative satisfying method. Then it may be necessary to examine in more detail the possible tradeoffs among the competing alternatives.

9.5.5 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 the particular values of these multiple objectives. Then, from among the

Fig. 9.15 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

non-inferior plans that satisfy 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.

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 type of fruit 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 be impacted by the plan.

9.5.6   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 let Z_1^B and Z_2^B be the values of the two respective objectives for plan *B*. Comparing both plans when a different objective is better 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 in this two-objective example. 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 (equally happy or satisfied) 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^*). 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).

$$(Z_1, Z_2^*) \approx (Z_1^A, Z_2^A). \qquad (9.16)$$

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.

$$(Z_1', Z_2^*) \approx (Z_1^B, Z_2^B) \qquad (9.17)$$

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 z_1^i estimated for each alternative plan i such that one is indifferent between (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 z_1^i will no doubt be greater than Z_3^i. Conversely, if $Z_3^* > Z_3^i$, then 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. The focus now is on the tradeoff between the values of objectives 1 and 2 given the same Z_3^*. A value $zz_1^{i'}$ must be selected such that the participants are indifferent between (z_1^i, Z_2^*, Z_3^*) and $(z\underline{z}_1^i, Z_2^*, Z_3^*)$. 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

$$(Z_1^i, Z_2^i, Z_3^i). \approx (z_1^i, Z_2^i, Z_3^*) \qquad (9.18)$$

In the next step involving the hypothetical plans just defined objective 3 remained constant and only objectives 1 and 2 were compared to get

$$(\underline{z}_1^i, Z_2^i, Z_3^*) \approx (zz_1^i, Z_2^*, Z_3^*) \qquad (9.19)$$

Hence

$$(Z_1^i, Z_2^i, Z_3^i) \approx (z_1^i, Z_2^i, Z_3^*) \approx (zz_1^i, Z_2^*, Z_3^*) \qquad (9.20)$$

Having done this for all n plans, there are now n hypothetical plans (zz_1^i, Z_2^*, Z_3^*) that differ only in the value of zz_1^i. All n plans can be ranked just based on the value of this single objective.

Each of these plan selection techniques requires the prior identification of discrete alternative plans.

9.5.7 Goal Attainment

The goal attainment method combines some of the advantages of both the weighting and constraint plan generation methods already discussed. 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 objective value, $Z_j(X)$, from its specified target, T_j. The problem is to

$$\text{minimize } D \quad (9.21)$$

subject to

$$g_i(X) = b_i \quad i = 1, 2, \ldots, m \quad (9.22)$$

$$w_j[T_j - Z_j(X)] \le D \quad j = 1, 2, \ldots, J \quad (9.23)$$

Constraints 9.22 contain the relationships among the decision variables in the vector X. They define the feasible region of decision variable values.

This method of multicriteria analysis can generate efficient or non-inferior plans by adjusting the weights and targets. It is illustrated for a two-objective problem in Fig. 9.16.

If the weights are equal, then the deviations will be equal and the resulting feasible solution will be the closest to the ideal but infeasible one. Unless $T_j \ge Z_j(X)$ some plans generated from a goal attainment method may be inferior with respect to the objectives being considered.

9.5.8 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

$$\text{minimize } \sum_j L_j(D_j) \quad (9.24)$$
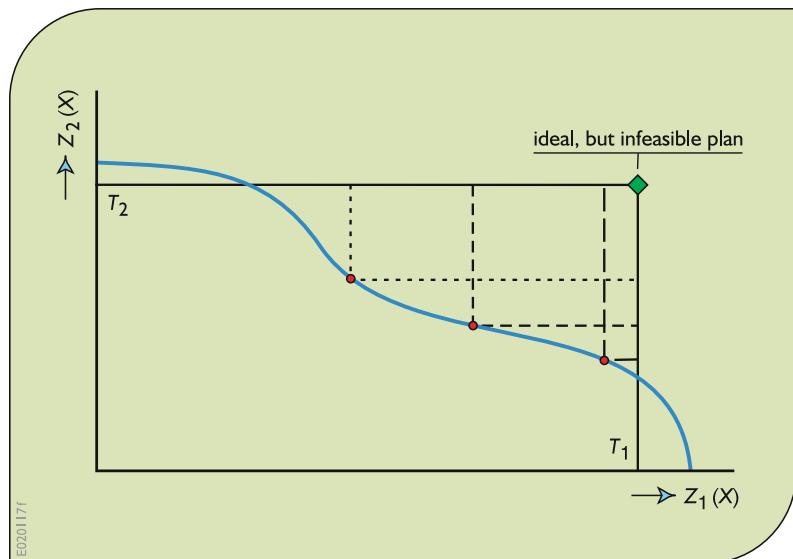
subject to:

$$g_i(X) = b_i \quad i = 1, 2, \ldots, m \quad (9.25)$$

$$T_j - Z_j(X) \le D_j \quad j = 1, 2, \ldots, J \quad (9.26)$$

If the loss functions are linear or piecewise linear the model can be solved using linear programming methods. Again, the target and loss values can be changed to generate alternative plans X.

Fig. 9.16 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

9.5.9 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 iterative approach, called the step method, requires, at each iteration preference, information from the participants in the process. 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 based on the relative range of values each objective j can assume, and on whether or not the participants have indicated satisfaction regarding a particular objective value obtained from a previous solution. If they are satisfied with the value of, say, an objective $Z_j(X)$, they must indicate how much of that value they would be willing to give up to obtain unspecified improvements in objectives whose values they consider unsatisfactory. This defines a lower bound on $Z_j(X)$. Then the weight w_j for that objective is set to 0, and the weights of all remaining objectives are recalculated. 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 Fig. 9.17. To work, they require the participation of the participants in the planning and management process.
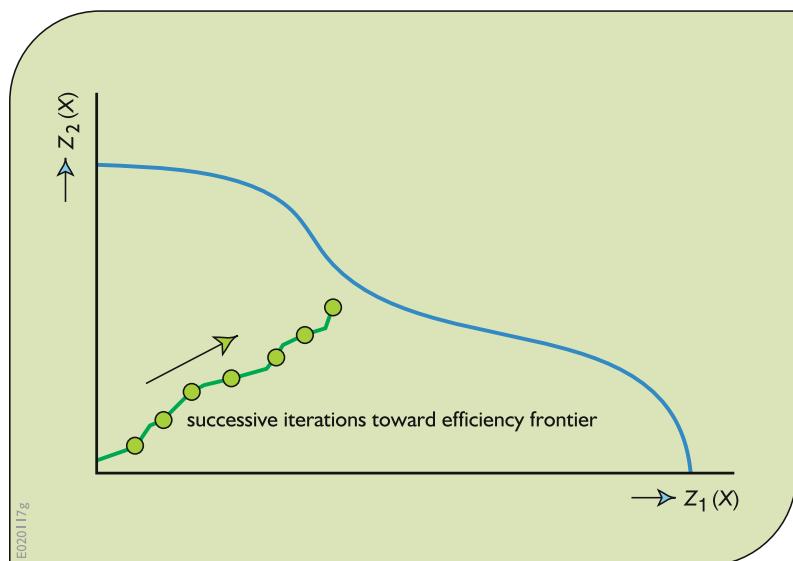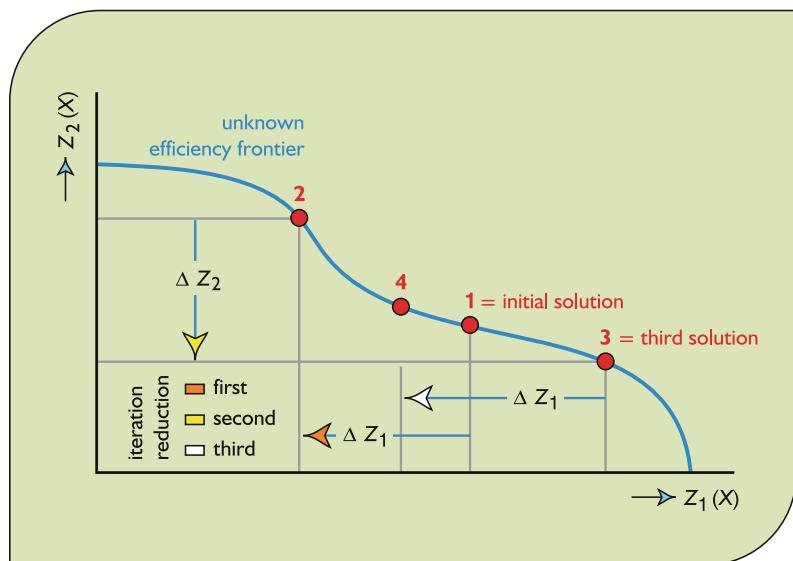
9.5.10 Plan Simulation and Evaluation

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 multiobjective planning (Cohon 1978; Steuer 1986). Most have been formulated in an optimization framework. This section describes ways of evaluating alternative water management plans or policies based on the time series of performance criteria values derived from simulation models.
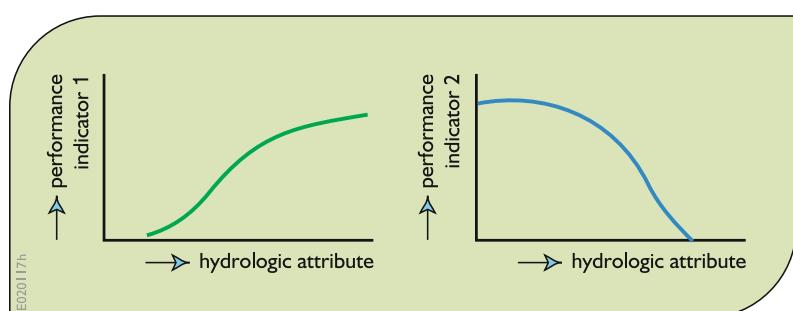
Simulation models of water resource systems yield time series of output variable values. These values in turn impact multiple system performance criteria, each pertaining to a specific interest and measured in its appropriate units. A process for evaluating alternative water resource management plans or policies based on these simulation model results includes the following steps:
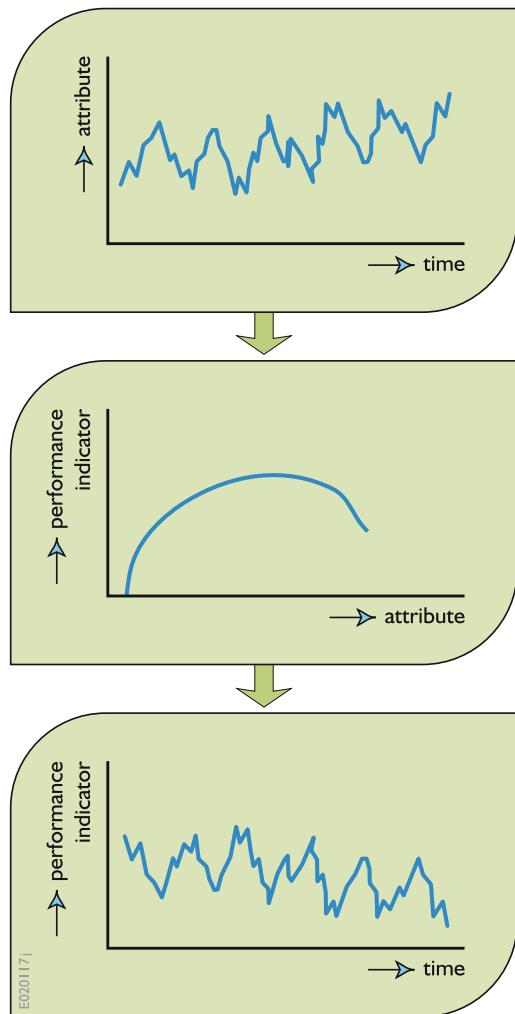
1. Identify system performance indicators that are impacted by one or more hydrologic attributes whose values will vary depending on the management policy or plan being simulated. For example, navigation benefits, measured in monetary units, might depend on water depths and velocities. Hydropower production is affected by water heads and

Fig. 9.17 Two interactive iterative multiple criteria approaches for identifying the tradeoffs of interest and possibly the best decision

Fig. 9.18 Performance indicators expressed as functions of simulated hydrological attributes
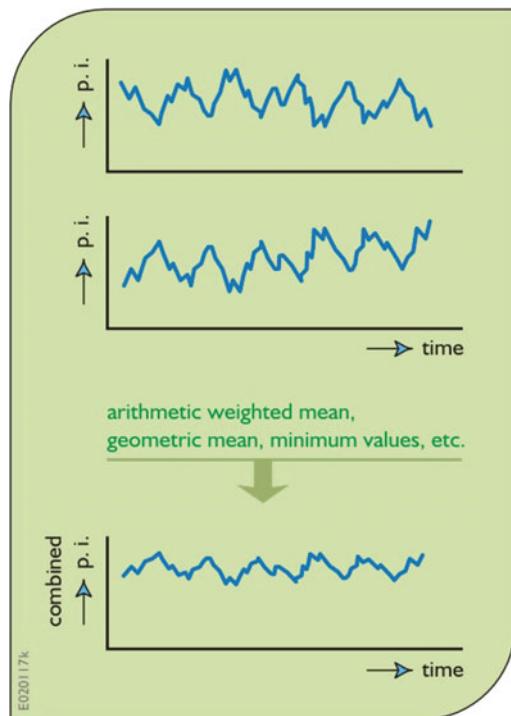
Fig. 9.19 Mapping a time series of hydrological attribute values into a corresponding time series of performance indicator values

Fig. 9.20 Combining multiple time series of values of a specified performance indictor into a single time series of values of that performance indicator

discharges through the power plant. Water quality might be expressed as the average or maximum concentration of various potential pollutants over a fixed time period at certain locations, and will depend in part on the flows. Ecological habitats may be impacted 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 hydrologic attributes. Figure 9.18 illustrates such functions. The units on each axis may differ for each such function.

3. Simulate to obtain time-series of hydrologic attribute values and map them into a time series of performance indicator values using the functional relationships defined in step 2. This step is illustrated in Fig. 9.19.

4. Combine multiple time series values for the same performance criterion, as applicable, as shown in Fig. 9.20. 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 impact 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 modeler 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 color-coded ranges and display on maps or in scorecards, as illustrated in Fig. 9.21.

The area under each exceedance curve is the mean. Different exceedance functions will result from different water management policies, as illustrated in Fig. 9.22.

One can establish thresholds to identify discrete zones of performance indicator values and assign a color 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 colored green; the worst value for each indicator can be colored red. The water management alternative having the most number of green boxes will stand out and will probably be considered more seriously than the alternative having the most 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, ecological habitats. The performance measures for each of these interests were identified and expressed as functions of one or more hydrologic 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.
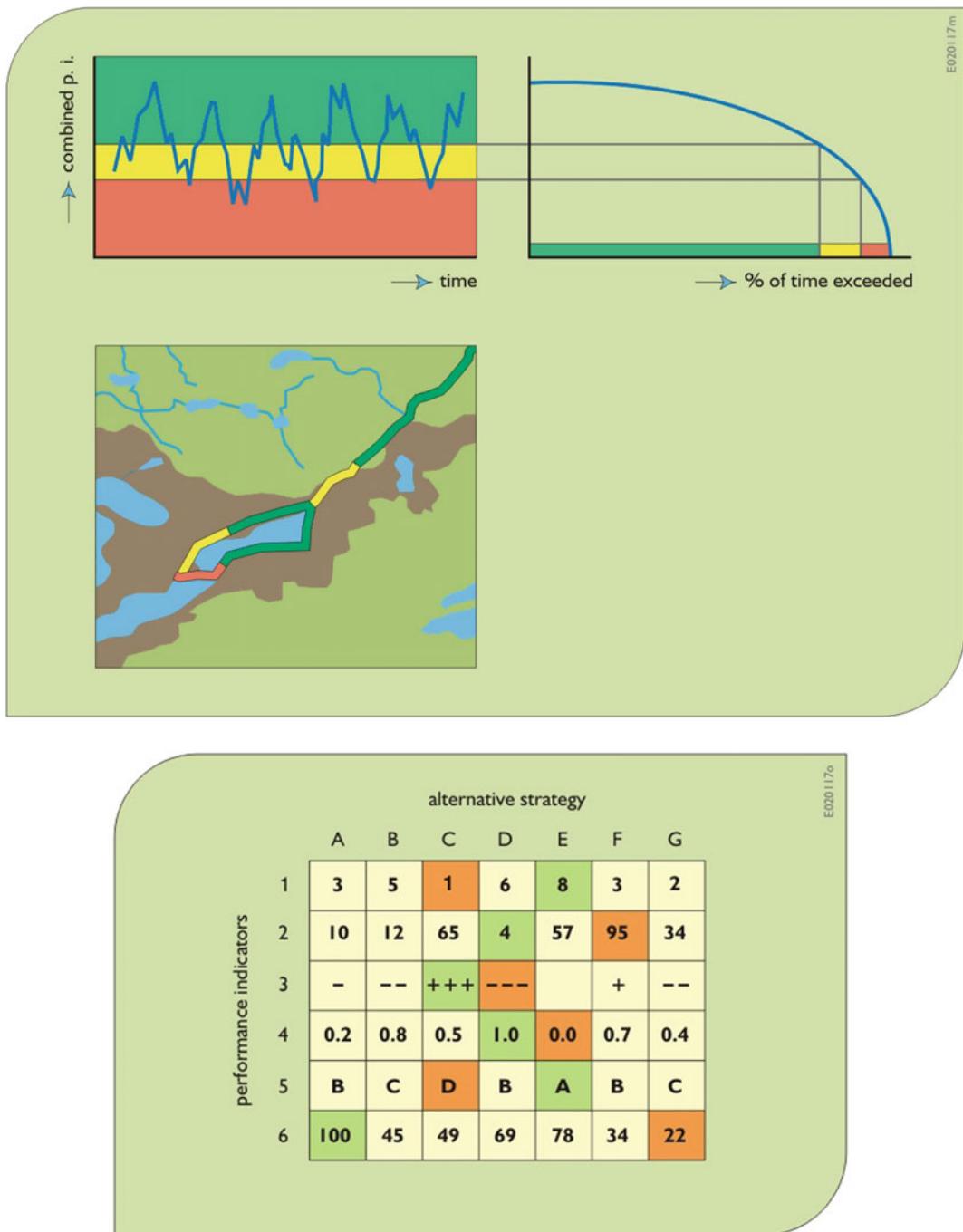
The same five-step process has been implemented in the Everglades restoration project in southern Florida. The Everglades is a long very wide and extremely flat 'river of grass' flowing generally south into, eventually, the Atlantic Ocean and Gulf of Mexico. This ecosystem restoration project has involved numerous local, state and federal agencies. The project impacts 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 billon. 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 in south Florida is 'to get the water right.' Those who manage the region's water are attempting to restore the ecosystem by restoring the hydrologic regime, i.e., the flows, depths, hydroperiods, and water qualities, throughout this region to what they think existed some 60 years ago. The trick is to accomplish this and still meet water supply, flood protection, and land development needs of those who live in the region. Clearly achieving a hydrologic 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? In either case what will be its likely impact on the ecosystem? Are there opportunities for changing hydrology to improve ecology? Where?

To address questions such as these in the Everglades, at least in a preliminary way prior to when more detailed ecological models will become available, this five-step approach outlined above is being applied. It is being used to extend their simulated hydrological predictions to produce relative values of ecological habitat suitability indicators for selected indicator species, as illustrated in Fig. 9.8, and topographic characteristics.
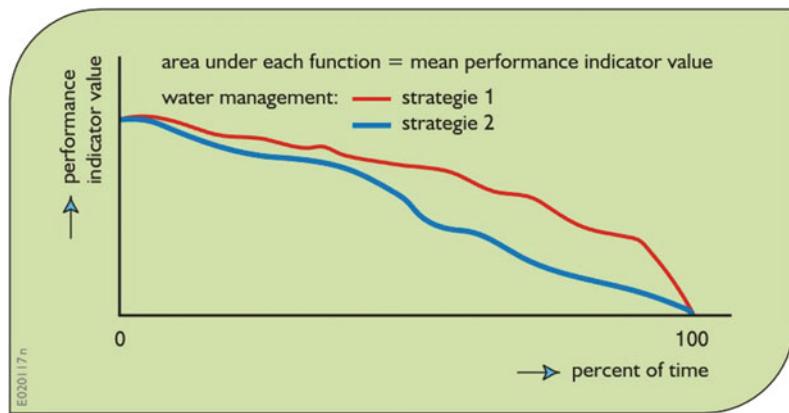
This five-step procedure does not find an 'optimal' water management policy. It can however contribute useful information to the political debate that must take place in the search of that

Fig. 9.21 Ways of summarizing and displaying time series performance indicator data involving exceedance distributions, and color-coded maps and scorecards. Color-coded map displays on computers can be dynamic, showing changes over time. *Green* and *red* colored scorecard entries indicate best and worst plan or strategy, respectively, for associated performance indicator

Fig. 9.22 Two exceedance functions showing decrease in mean performance value if policy 2 is followed compared to Policy 1

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 understand just how this multiobjective 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 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.

9.6 Statistical Summaries of Performance Criteria

There are numerous ways of summarizing time series data in addition to the methods just mentioned above. 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 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
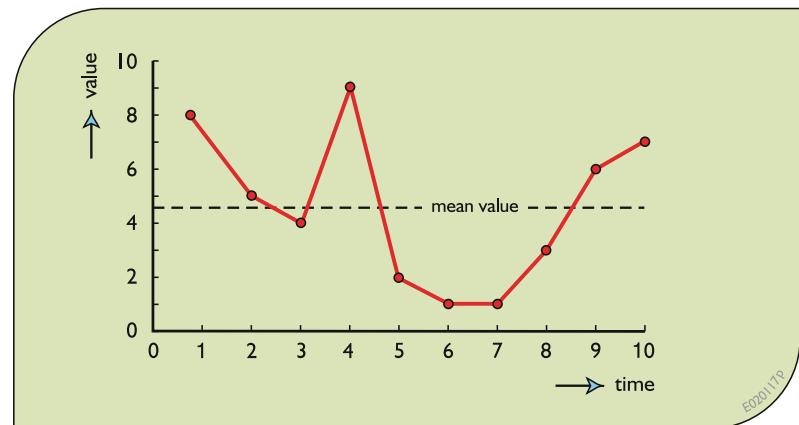
$$
\sum_t^T (X_t - \underline{X})^2 / T = \\
\left[(8-4.6)^2 + (5-4.6)^2 + \ldots + (7-4.6)^2 \right] / 10 \\
= 7.44
$$

(9.27)

A plot of these values and their mean is shown in Fig. 9.23. The mean and variance for the time series shown in Fig. 9.23 however are the same for its upside down image, as shown in Fig. 9.24. They do not even depend on the order of the time series data.
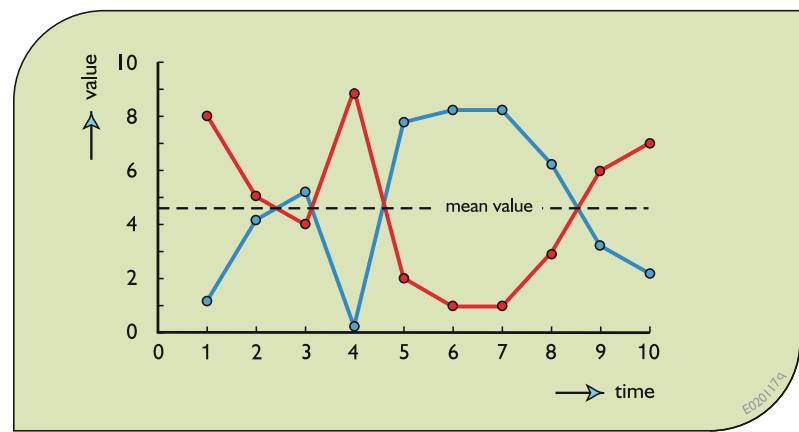
Consider these two sets of time series shown again in Fig. 9.25, each having the same mean and variance. Assume that any value equal or less than the dashed line (just above 2) is considered unsatisfactory. This value is called a threshold value, dividing the time series data into satisfactory and unsatisfactory values.

It is clear from Fig. 9.25 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 time series shown in blue. However, the
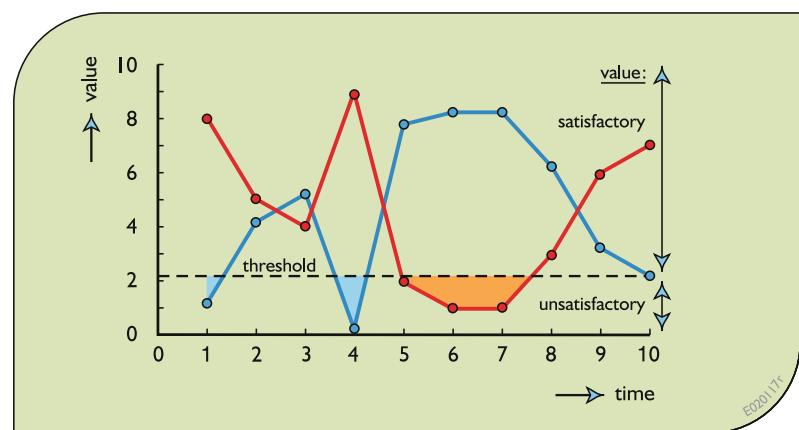
Fig. 9.23 Plot of time series data having a mean value of 4.6 and a variance of 7.44

Fig. 9.24 A plot of two different time series having the same mean and variance

Fig. 9.25 Threshold value distinguishing values considered satisfactory, and those considered unsatisfactory

maximum extent of failure when it, the red series, failed was less than the blue time series. These characteristics can be captured by the measures of reliability, resilience and vulnerability (Hashimoto et al. 1982).

9.6.1 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

$$\text{Reliabilty}[X] = [\text{number of time periods } t \text{ such that } X_t \geq X^T]/n \tag{9.28}$$

The reliability of the red time series is 0.7. It failed three times in 10. The reliability of the blue time series is also 0.7, failing three times in 10.

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 preferred to one whose reliability is much higher, but when a failure does occur, it is likely to be much more severe. Resilience and vulnerability measures can provide measures of these system characteristics.

9.6.2 Resilience

Resilience can be expressed as the probability that if 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

$$\text{Resilience}[X] = [\text{number of times a satisfactory value follows an unsatisfactory value}]/[\text{number of times an unsatisfactory value occurred}] \tag{9.29}$$

Resilience is not defined if no unsatisfactory values occur in the time series. For the time series shown in red, the resilience is 1/3, again assuming the value of 2 or less is considered a failure. For the time series shown in blue the resilience is 2/2 = 1. We cannot judge the resilience of the blue time series based on the last failure in period 10 because we do not have an observation in period 11.

9.6.3 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

$$\text{Vulnerability}[X] = [\text{sum of positive values of } (X^T - X_t)] / [\text{number of times an unsatisfactory value occurred}] \tag{9.30}$$

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 blue line in Fig. 9.25 is $[(2 - 1.2) + (2 - 0.2)]/2 = 1.3$.

So, while in this example the reliability of red time series equals that of the blue time series, the resilience of the blue time series is better than that of the red time series. Yet the expected vulnerability of the red time series is less than that of the rotated blue time series. This shows the typical tradeoffs one can observe using these three measures of system performance.

9.7   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 resource 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 multiobjective 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 tradeoffs among the interests of multiple 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 that the "analyst" make these tradeoffs and just 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 expert. However 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 further development and use of practical analytical multiobjective 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 dictate substantially 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, a reminder that the decisions being made at the current time are only those in a sequence of decisions that will continue to 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 and sustainability (ASCE 1999).

References

American Society of Civil Engineers (ASCE). (1999). Sustainability criteria for water resource systems, ASCE Press, Reston, Va. (also in D. P. Loucks & J. S. Gladwell (Eds.), 2000. Cambridge, UK: Cambridge University Press.

Cohon, J. L. (1978). *Multiobjective programming and planning*. New York: Academic Press.

Gregory, R. S., & Keeney, R. L. (1994). Creating policy alternatives using stakeholder values. *Management Science, 40*, 1035–1048.

Hammond, J. S., Keeney, R. L., & Raiffa, H. (1999). *Smart choices: A practical guide to making better decisions*. Boston, Mass: Harvard Business School Press.

Hipel, K. W., Fang, L., Cullman, J., & Bristow, M. (Eds.). (2015). *Conflict resolution in water resources and environmental management, conflict resolution in water resources and environmental management*. Heidelberg, New York, Dordrecht, London: Springer Cham.

Madani, K. (2010). Game theory and water resources. *Journal of Hydrology, 381*(3), 225–238.

Madani, K., & Hipel, K. W. (2011). Non-cooperative stability definitions for strategic analysis of generic water resources conflicts. *Water Resources Management, 25*(8), 1949–1977.

Steuer, R. E. (1986). *Multiple criteria optimization: Theory, computation, and application*. New York: Wiley.

Additional References (Further Reading)

Chen, H. W., & Chang, N.-B. (1998). Water pollution control in the river basin by fuzzy genetic algorithm-based multiobjective programming modeling. *Water Science and Technology, 37*(8), 55–63.

Cieniawski, S. E., Eheart, J. W., & Ranjithan, S. (1995). Using genetic algorithms to solve a multiobjective groundwater monitoring problem. *Water Resources Research, 31*(2), 399–409.

Eschenauer, H., Koski, J., & Osyczka, A. (1990). *Multicriteria design optimization*. Berlin: Springer-Verlag.

Gupta, H. V., Sorroshian, S., & Yapo, P. O. (1998). Towards improved calibration of hydrologic models: Multiple and non-commensurable measures of information. *Water Resources Research, 34*(4), 751–763.

Haimes, Y. Y., Hall, W. A., & Freedman, H. T. (1975). *Multiobjective optimization in water resources systems*. Amsterdam: Elsevier Scientific.

Halhal, D., Walters, G. A., Ouazar, D., & Savic, D. A. (1997). Multi-objective improvement of water distribution systems using a structure messy genetic algorithm approach. *Journal of Water Resources Planning and Management ASCE, 123*(3), 137–146.

Hashimoto, T., Stedinger, J. R., & Loucks, D. P. (1982a). Reliability resiliency and vulnerability criteria for water resource system performance evaluation. *Water Resources Research, 18*(1), 14–20.

Hashimoto, T., Loucks, D. P., & Stedinger, J. R. (1982b). Robustness of water resource systems. *Water Resources Research, 18*(1), 21–26.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley.

Loucks, D. P., Stedinger, J. R., & Haith, D. A. (1981). *Water resource systems planning and analysis*. Englewood Cliffs, NJ: Prentice Hall.

Maas, A., et al. (1962). *Design of water resource systems*. Cambridge: Harvard University Press.

Reed, P. M., Minsker, B. S., & 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, 3*(2), 71–89. April.

Ritzel, B. J., Eheart, J. W., & Ranjithan, S. (1994). Using genetic algorithms to solve a multiple objective groundwater pollution containment problem. *Water Resources Research, 30*(5), 1589–1603. May.

Sawaragi, Y., Nakayama, H., & Tanino, T. (1985). *Theory of multiobjective optimization*. Orlando, Florida: Academic Press Inc.

Statnikov, R. B., & Matusov, J. B. (1995). *Multicriteria optimization and engineering*. New York: Chapman and Hall.

Yeh, C.-H., & Labadie, J. W. (1997). Multiobjective watershed-level planning of storm-water detention systems. *Journal of Water Resources Planning and Management, 123*(6), 336–343. November/December.

Zeleny, M. (1982). *Multiple criteria decision making*. New York: McGraw-Hill.

Exercises

9.1 Distinguish between multiple purposes and multiple objectives and give some examples of complementary and conflicting purposes and objectives of water resources projects.

9.2 Assume that farmers' demand for water q is a linear function $a - b(p)$ of the price p, where $a, b > 0$. Calculate the farmers' willingness to pay for a quantity of water q. If the cost of delivering a quantity of water q is cq, $c > 0$, how much water should a public agency supply to maximize willingness to pay minus total cost? If the

local water district is owned and operated by a private firm whose objective is to maximize profit, how much water would they supply and how much would they earn? The farmers' consumer surplus is their willingness to pay minus what they must pay for the resource. Compare the farmers' consumer surplus in two cases. Do the farmers loose more than the private firm gains by moving from the social optimum to the point that maximizes the firm's profit? Illustrate these relationships with a graph showing the demand curve and the unit cost c of water. Which areas on the graph represent the firm's profits and the farmers' consumer surplus?

9.3   Consider the water allocation problem used in the earlier chapters of this book. The returns, $B_i(X_i)$ from allocating X_i amount of water to each of three uses i are as follows, along with the optimal allocations from the point of view of each use.

$$B_1(X_1) = 6X_1 - X_1^2 \quad \rightarrow \quad X_1^{opt} = 3$$
$$\text{and} \quad B_1^{max} = B_1(X_1^{opt}) = 9$$
$$B_2(X_2) = 7X_2 - 1.5X_2^2 \quad \rightarrow \quad X_2^{opt} = 7/3$$
$$\text{and} \quad B_2^{max} = B_2(X_2^{opt}) = 147/18$$
$$B_3(X_3) = 8X_3 - 0.5X_3^2 \quad \rightarrow \quad X_3^{opt} = 8$$
$$\text{and} \quad B_3^{max} = B_3(X_3^{opt}) = 32$$

Consider this a multiobjective problem. Instead of finding the best overall allocation that maximizes the total return assume the objectives are to maximize the returns from each user.

Show how the weighting, constraint, goal attainment, and goal programming methods can be used to identify the tradeoffs among each of the three objectives for any limiting total amount of water, for example, 6.

9.4   Under what circumstances will the weighting and constraint methods fail to identify efficient solutions?

9.5   A reservoir is planned for irrigation and low flow augmentation for water quality control. A storage volume of 6×10^6 m^3 will be available for those two conflicting uses each

year. The maximum irrigation demand (capacity) is 4×10^6 m^3. Let X_1 be the allocation of water to irrigation and X_2 the allocation for downstream flow augmentation. Assume that there are two objectives, expressed as

$$Z_1 = 4X_1 - X_2$$
$$Z_2 = -2X_1 + 6X_2$$
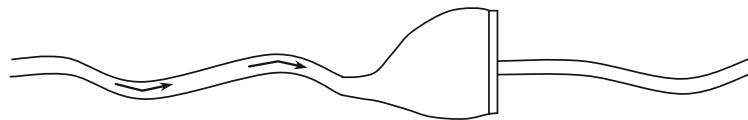
(a) Write the multiobjective planning model using a weighing approach and a constraint approach.
(b) Define the efficient frontier. This requires a plot of the feasible combinations of X_1 and X_2.
(c) Assume that various values are assigned to a weight W_1 for Z_1 whereas weight W_2 for Z_2 is constant and equal to 1, verify the following solutions to the weighing model.

W_1	X_1	X_2	Z_1	Z_2
>6	4	0	16	−8
6	4	0 to 12	16 to 14	−8 to 4
<6 to >1.6	4	2	14	4
1.6	4 to 0	2 to 6	14 to −6	4 to 36
<1.6	0	6	−6	36
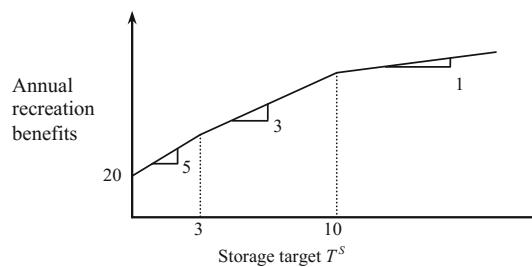
9.6   Show that the following benefit, loss, and cost functions can be included in a linear optimization problem for finding the active storage volume target T^s, annual release target T^R and the actual storage releases R_t in each within-year period t, and the reservoir capacity K. The objective is to maximize annual net benefits from the construction and operation of the reservoir. Assume that the inflows are known in each of 12 within-year periods t. Note that the loss function associated with reservoir recreation is independent of the value of T^s,

unlike the loss function associated with reservoir releases. Structure the complete linear programming model. Define all variables used that are not defined below. Let $\delta_t T^R$ be the known release target in period t.
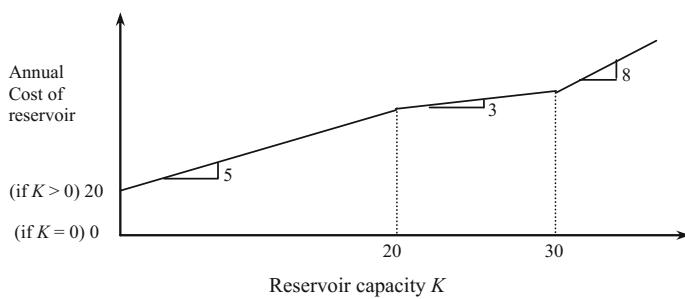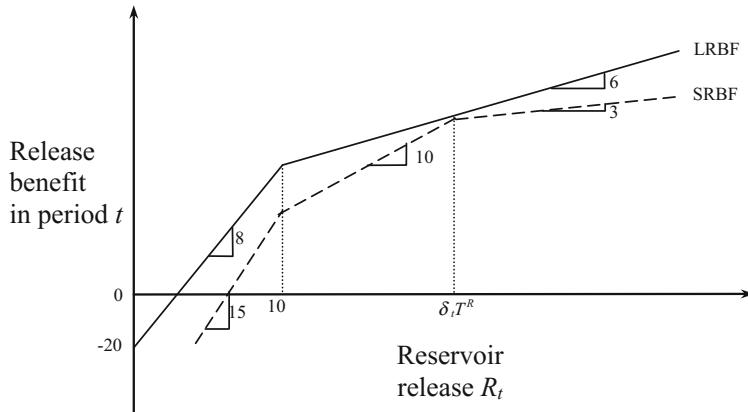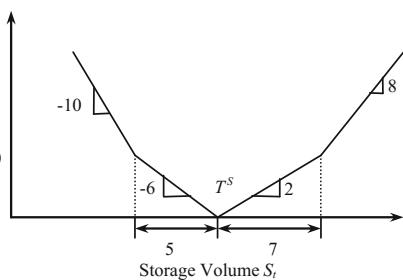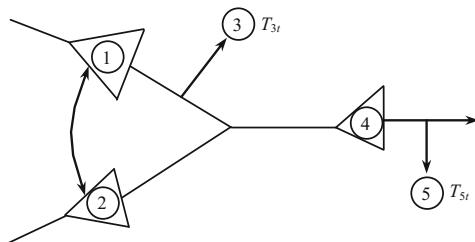
S_t = initial storage volume
K = reservoir

Annual recreation benefits

20

5

3

3

1

3   10

Storage target T^S

Recreation Loss in periods t= 6 through 9 (0 otherwise)

-10

-6

T^S

2

8

5   7

Storage Volume S_t

Release benefit in period t

LRBF

SRBF

6

3

10

8

0

15   10

$\delta_t T^R$

-20

Reservoir release R_t

Annual Cost of reservoir

(if $K > 0$) 20

(if $K = 0$) 0

5

3

8

20   30

Reservoir capacity K

9.7 For the river basin shown, potential reservoirs exist at sites $i = 1, 2$, and 4 and a diversion can be constructed between sites 1 and 2. The cost $C_i(K_i)$ of each reservoir i is a function of its active storage capacity K_i. The cost of the diversion canal is $C_i(Q_i)$ where Q is the flow capacity of the canal. The cost of diverting a flow Q_{ijt} from site i to site j is $C_{ij}(Q_{ijt})$. The two users at sites 3 and 5 have known target allocations (demands) T_{it} in each period t. The return flow from use 3 is 40 % of that allocated to use 3. Construct a model for finding the least cost of meeting various percentages of the target demands. Assume that the natural streamflows Q_t^i at each site i in each period t, are known.

9.8 Suppose that there exist two polluters, A and B, who can provide additional treatment, X_A and X_B, at a cost of $C_A(X_A)$ and $C_B(X_B)$, respectively. Let W_A and W_B be the waste produced at sites A and B, and $W_A(1 - X_A)$ and $W_B(1 - X_B)$ be the resulting waste discharges at site A and B. These discharges must be no greater than the effluent standards E_A^{max} and E_B^{max}. The resulting pollution concentration $a_{Aj}(W_A(1 - X_A)) + a_{Bj}(W_B(1 - X_B)) + q_j$ at various sites j must not exceed the stream standards S_j^{max}. Assume that total cost and cost inequity [i.e., $C_A(X_A) + C_B(X_B)$ and| $C_A(X_A) - C_B(X_B)$] are management objectives to be determined

(a) Discuss how you would model this multiobjective problem using the weighting and constraint (or target) approaches.

(b) Discuss how you would use the model to identify efficient, non-inferior (Pareto-optimal) solutions.

(c) Effluent standards at sites A and B and ambient stream standards at sites j could be replaced by other planning objectives (e.g., the minimization of waste discharged into the stream). What would these objectives be, and how could they be included in the multiobjective model?

9.9 (a) What conditions must apply if the goal attainment method is to produce only non-inferior alternatives for each assumed target T_k and weight w_k?

(b) Convert the goal programming objective deviation components $w_i \, (z_i^* - z_i(\bar{x}))$ to a form suitable for solution by linear programming.

9.10 Water quality objectives are sometimes difficult to quantify. Various attempts have been made to include the many aspects of water quality in single water quality indices. One such index was proposed by Dinius (Social Accounting Systems for Evaluating Water Resources, Water Resources Research, Vol. 8, 1972. pp. 1159–1177). Water quality, Q, measured in percent is given by

$$Q = \frac{w_1 Q_1 + w_2 Q_2 + \ldots + w_n Q_n}{w_1 + w_2 + \ldots + w_n}$$

where Q_i is the ith quality constituent (dissolved oxygen, chlorides, etc.) and w_i is the weight or relative importance of the ith quality constituent. Write a critique on the use of such an index in multiobjective water resources planning.

9.11 Let objective $Z_1(X) = 5X_1{}_{-2}X_2$ and objective $Z_2(X) = -X_1 + 4 \, X_2$. Both are to be maximized. Assume that the constraints on variables X_1 and X_2 are:

1. $-X_1 + X_2 \leq 3$
2. $X_1 \leq 6$
3. $X_1 + X_2 \leq 8$
4. $X_2 \leq 4$
5. $X_1, X_2 \geq 0$

(a) Graph the Pareto-optimal or non-inferior solutions in decision space.
(b) Graph the efficient combination of Z_1 and Z_2 in objective space.
(c) Reformulate the problem to illustrate the weighting method for defining all efficient solutions of part (a) and illustrate this method in decision and objective space.
(d) Reformulate the problem to illustrate the constraint method of defining all efficient solutions of part (a) and illustrate this method in decision and objective space.
(e) Solve for the compromise set of solutions using compromise programming as defined by

$$\text{Minimize } \left[w_1 \left(Z_1^* - Z_1 \right)^\alpha + w_2 \left(Z_2^* - Z_2 \right)^\alpha \right]^{1/\alpha}$$

where Z_i^* represents the best value of objective i with all weights w equal to 1 and α equal to 1, 2, and ∞.

9.12 Illustrate the procedure for selecting among three plans, each having three objectives, using indifference analysis. Let Z_{ji} represent the value of objective i for plan j. The values of each objective for each plan are given below. Assume that each objective is to be maximized. Assume that an identical indifference function for all trade-offs between pairs of objectives, namely one that implies you are willing to give up twice as many units of your higher (larger) objective value to gain one unit of your lower (smaller) objective value. [For example, you would be indifferent to two plans having as their three objective values (30, 5, 10) and (20, 5, 15).] Rank these three plans in order of preference.

Plan 1 : (5, 25, 15);
Plan 2 : (10, 20, 10);
Plan 3 : (15, 10, 15)

Water Quality Modeling and Prediction

<div style="text-align:right">

10

</div>

The most fundamental human needs for water are for drinking, cooking, and personal hygiene. The quality of the water used to meet these needs must pose no risk to human health. The quality of the water in nature also impacts the condition of all living organisms found in aquatic ecosystems that we depend upon for our own wellbeing. At the same time watersheds and their water bodies serve as convenient sinks for domestic, industrial, and agricultural wastes. Runoff from agricultural and urban lands containing excess nutrients, oils, and solid wastes together with direct point source discharges of wastewaters into water bodies 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 can assist in predicting the water quality impacts of alternative land and water management policies and practices. This chapter introduces some approaches to water quality modeling, leaving descriptions of more advanced methods to textbooks devoted solely to this subject.

10.1 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, water users must then either pay for water treatment or incur increased risks of using lower quality water. As populations and economies grow, more wastewater pollutants are generated. Many of these are discharged into surface and ground water bodies. Increasingly the major efforts and costs involved in water management are aimed at water quality protection and management. Conflicts among various users of water are increasingly over issues involving water quality.

Natural water bodies are able to serve many uses. One of them is the transport and assimilation of many waterborne wastes. As natural water bodies transport and assimilate wastes, their quality changes. If the quality of water drops to the extent that other uses are adversely impacted, the assimilative capacities of those water bodies have been exceeded with respect to those impacted uses. Water management measures are actions taken to ensure that the total pollutant loads discharged into receiving water bodies do not exceed the waste assimilative capacity of those water bodies and that the quality meets the quality standards set for those waters.

What uses depend on water quality? Almost all one can identify. As everyone knows, all living organisms require water of sufficient quantity and quality to survive. Different aquatic species can tolerate different levels of pollutant concentrations that impact water quality. In much of the developed world it is no longer "safe" to drink natural surface or ground waters. Treatment is usually required before these waters are safe for humans to drink. Treatment is not a practical option for improving the quality of water found in

nature yet this is the water that impacts the health of fish and shellfish and other organisms in natural aquatic ecosystems. Hence the focus in practice is on the use of wastewater treatment facilities to improve the quality of effluents being discharged into natural water bodies.

Standards specifying minimum acceptable levels of quality are commonly set for most ambient waters. Various uses may have their own quality requirements as well. Irrigation water must not be too saline nor contain toxic substances that can be absorbed by the plants or destroy the microorganisms in the soil. Water quality standards for industry can be very demanding, depending on course of the requirements of particular industrial processes.

Domestic wasteloads can contain high concentrations of bacteria, viruses, and other organisms that impact human health. High organic loadings can reduce dissolved oxygen (DO) to levels that can kill parts of the aquatic ecosystem and cause obnoxious odors. Nutrient loadings from both urban and agricultural land runoff can cause excessive algae growth that in turn may degrade the water aesthetically, inhibit boating and swimming, and upon death cause low DO levels. Toxic heavy metals and other micropollutants 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 to non-point sources. A common approach to controlling point source discharges, such as 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. Nevertheless they are relatively easy to monitor and control.

Non-point sources such as agricultural runoff or atmospheric deposition are not as easily controlled and hence it is difficult to apply effluent standards to non-point source pollutants. Pollutant loadings from non-point sources can be much higher than point source loadings. Management of non-point water quality impacts requires a more ambient-focused water quality management program.

The goal of an ambient water quality management program 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 hydrologic, ecological, and land use conditions, the potential uses of the receiving water bodies, 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 be in the future, especially from area-wide non-point sources. In addition to uncertainties inherent in measuring water quality, there are uncertainties in models used 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 to the complexities of actual water systems. If water quality models are being used to inform those setting standards and permissible waste loadings, these limitations and uncertainties should be understood and addressed.

10.2   Establishing Ambient Water Quality Standards

A first step in setting water quality standards for a water body is to identify the intended uses of that water body, whether a lake, a section of a stream, or areas of an estuary. The most restrictive (in terms of water quality) 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 hydrologic and geomorphic changes that impact the quality of the water body.

The designated use dictates the appropriate type of water quality standard. For example, the standards of a water body whose designated use involves human contact recreation should protect humans from exposure to microbial pathogens while swimming, wading, or boating. Other uses might require standards to protect humans and aquatic life including fish, shellfish, and other wildlife from consuming harmful substances.

Standards set upstream may impact the uses of water downstream. For example, the water quality of small headwater streams may affect the ability of a downstream area to achieve a particular designated use such as being "fishable" or "swimmable." In this case, the designated use for the smaller upstream water body may be defined in terms of the achievement of the designated use of the larger downstream water body.

In many areas human activities have sufficiently altered the landscape and aquatic ecosystems to the point where they cannot be restored to their pre-disturbance condition. For example, a reproducing trout fishery in downtown Paris, Philadelphia, Phnom Penh, or Prague may be desired by some, but may not be attainable because of the development history of the areas or the altered hydrologic regimes of the rivers flowing through them. Similarly, health considerations would preclude designating an area for shellfish harvesting near the outfall of a sewage treatment plant. Ambient water quality standards must be realistic.
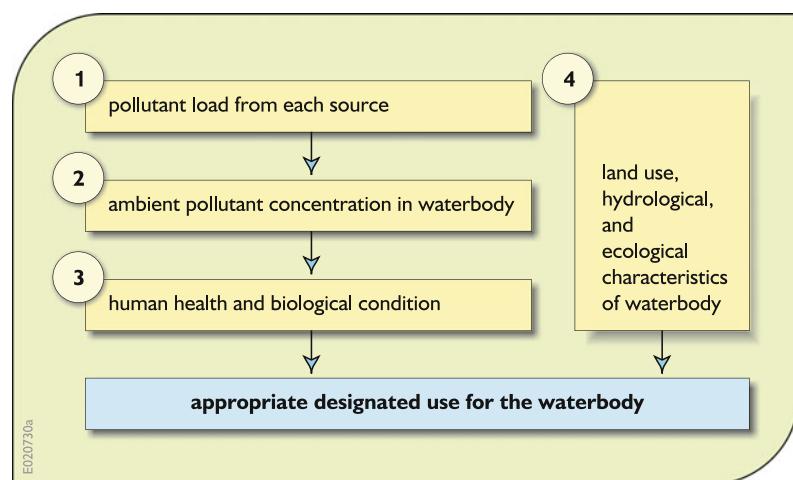
Decisions regarding the appropriate use for water bodies can be informed by the use of water quality prediction models. However, the final standard selection should reflect a social consensus made in consideration of the current condition of the watershed, its pre-disturbance condition, the advantages derived from a certain designated use, and the costs of achieving the designated use.

10.2.1   Water Use Criteria

The designated use is a qualitative description of a 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 Fig. 10.1.

Box 1 of Fig. 10.1 contains information about the pollutant discharges, e.g., from a treatment plant or in runoff (e.g., biological oxygen demand, ammonia (NH_3), pathogens, and suspended sediments). Effluent standards specifying maximum permissible loadings may apply to these pollutant loadings. Criteria in Boxes 2 and 3 are possible measures of ambient water quality conditions. Box 2 includes measures of a water

Fig. 10.1 Factors considered when determining designated use and associated water quality standards

quality parameter such as DO, pH, nitrogen concentration, suspended sediment, or temperature. Criteria closer to the designated use (e.g., Box 3) 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 hydrologic criterion), abundance of nonindigenous taxa, some quantification of channel modification (e.g., decrease in sinuosity), etc. (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 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, as previously discussed, the desired condition of a small headwater stream may need to be chosen as it relates to other water bodies downstream. Thus, an ambient nutrient criterion may be set in a small headwater stream to ensure a designated use downstream, even if there are no local adverse impacts resulting from the nutrients in the small headwater stream. 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.

10.3 Water Quality Model Use

Monitoring data are the preferred form of information for identifying impaired waters. Model predictions might be used in addition to or instead of monitoring data for several reasons

1. Modeling could be feasible in some situations where monitoring is not.
2. Integrated monitoring and modeling systems could provide better information than monitoring or modeling alone for the same total cost. For example, regression analyses that correlate pollutant concentration with some more easily measurable factor (e.g., streamflow) could be used to extend monitoring data for preliminary planning purposes. Models can also be used 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 Chaps. 7 and 9).
3. Modeling 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 begins operating, or the effect of increased industrial growth and effluent discharges.

A simple, but useful, modeling 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 stream flow 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 lead to 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 nonindigenous taxa, and over-harvesting of fishes. Indirect effects of humans include land cover changes that alter the rates of delivery of water, pollutants, and sediment to water bodies.

Direct and indirect environmental effects of human activities can include

- 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, and
- 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. The 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, models are simple qualitative conceptual representations of the relationships among important variables and indicators of those variables. 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.

10.3.1   Model Selection Criteria

There was a time when if one needed a water quality model, they had to build it. Today there exist a wide range of water quality models for various types of water bodies, and for various contaminants, and hence it makes little sense to build another one if an existing model will suffice. This section discusses criteria that can be used to select a particular model.

Water quality predictive models can include both mathematical expressions and expert scientific judgment. They may be process-based (mechanistic) models or data-based (statistical) models. Quality models used for planning and management should link management options to meaningful response variables (e.g., 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 alternative options. To do this requires prediction error estimates (Chap. 6).

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. Complex water quality problems 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. Finally, the cost of maintaining and updating the model during its use must be acceptable.

Although predictions are typically made using mathematical models, there are certainly situations where expert judgment can be just as good. Reliance on professional judgment and simpler models is often acceptable, especially when limited data exist.

Highly detailed models require more time and are more expensive to develop and apply. Complex modeling studies should be undertaken only if warranted by the complexity of the

management problem. More complex modeling will not necessarily assure that uncertainty is reduced, and in fact added complexity can compound problems of uncertainty analyses (Chap. 8).

Placing a priority on process description usually leads to complex mechanistic model development and use over simpler mechanistic or empirical models. In some cases this may result in unnecessarily costly analyses for effective decision-making. In addition, physical, chemical, and biological processes in terrestrial and aquatic environments are far too complex to be fully represented in even the most detailed models. For water quality management, the primary purpose of modeling should be to support decision-making. The inability to completely describe all relevant processes can be accounted for by quantifying the uncertainty in the model predictions.

10.3.2   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 limiting the nutrient loadings to the estuary that can promote algae growth and blooms, and subsequent DO deficits. The modeling 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 Fig. 10.2. 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 bottom water. Fish and shellfish may die or become weakened and more vulnerable to disease.
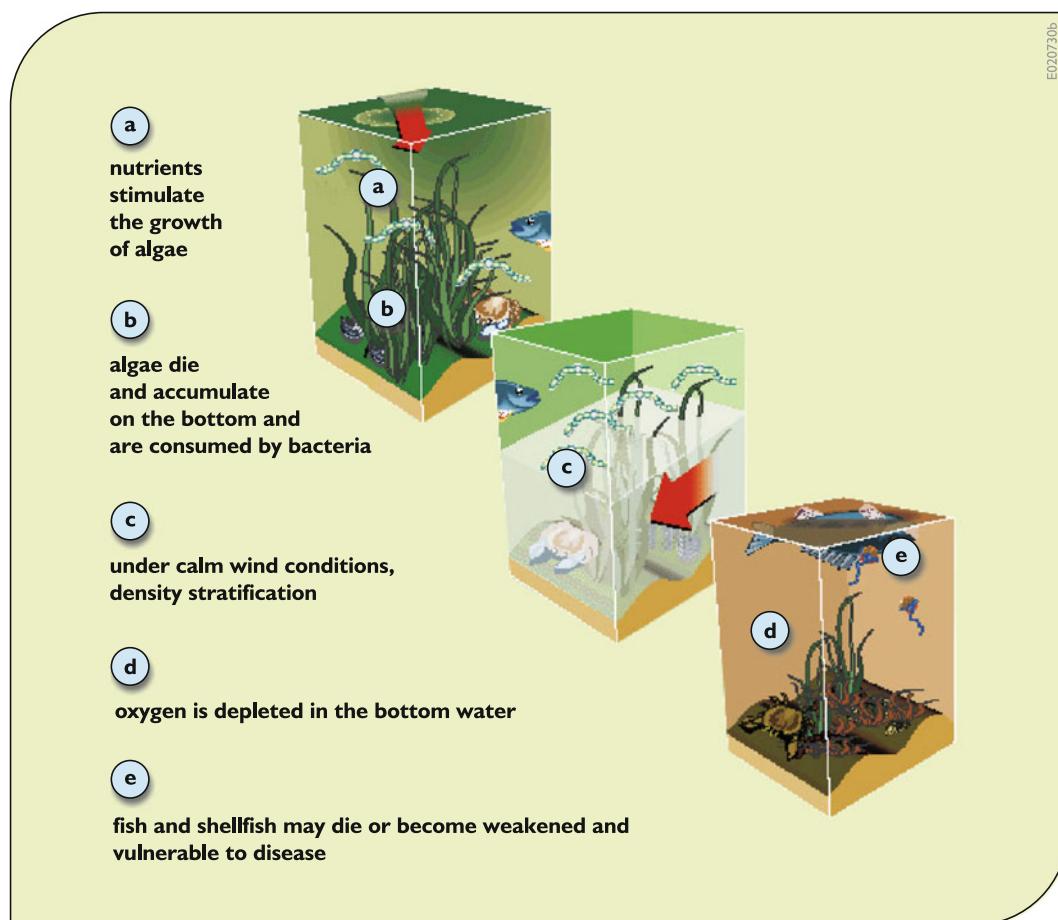
A model consisting of a sequence of conditional probabilities can be defined to predict the probability of shellfish and fish abundance based on upstream nutrient loadings causing problems with fish and shellfish populations into the estuary. These conditional probabilities can be judgmental, mechanistic, and/or statistical. Each conditional probability can be a separate submodel. Assuming each submodel can identify a conditional probability distribution, the probability $\Pr\{C|N\}$ of a specified amount of carbon, C, given some specified loading of a nutrient, say nitrogen, N, equals the probability $\Pr\{C|A\}$ of that given amount of carbon given a concentration of algae biomass, A, times the probability $\Pr\{A|N, R\}$ 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.

$$\Pr\{C|N\} = \Pr\{C|A\}\Pr\{A|N, R\}\Pr\{R\} \quad (10.1)$$

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 based on the river flow and nitrogen loadings. Similarly to predict the production of carbon based on 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 predict the duration of stratification and the frequency of hypoxia given both the stratification duration and sediment oxygen demand. Expert judgment 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

Fig. 10.2 The negative impacts of excessive nutrients in an estuary (NRC 2001)
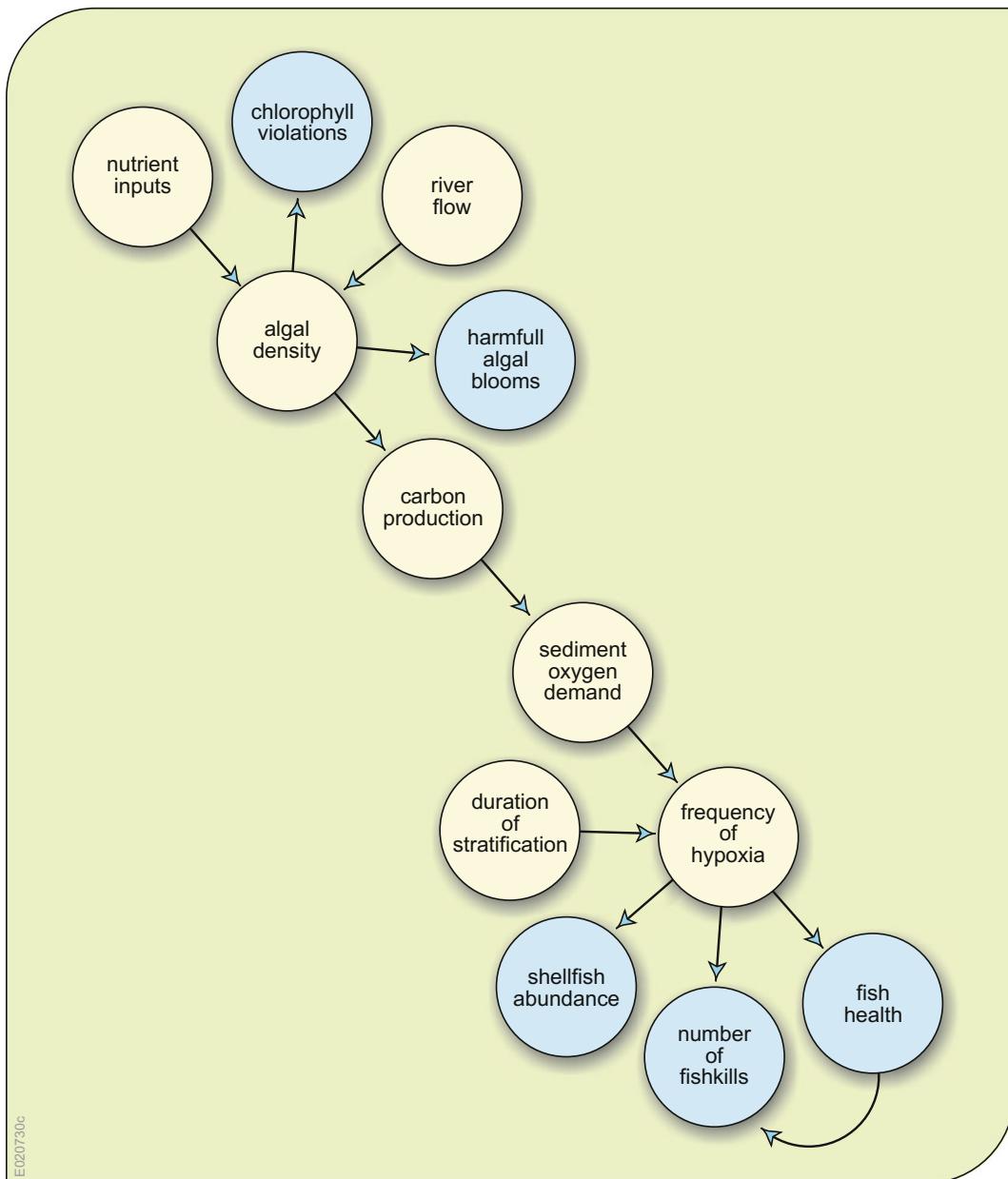
the network in Fig. 10.3 can reflect a combination of simple mechanisms, statistical (regression) fitting, and expert judgment.

Advances in mechanistic modeling of aquatic ecosystems have resulted in our ability to include greater process (especially trophic) detail and complexity, as well as to perform dynamic simulations. Still, 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 modeling (Chapra and Reckhow 1983; Chapra 1997; McCutcheon 1989; Thomann and Mueller 1987; Orlob 1983; Schnoor 1996) as well as the current literature.

10.3.3 Model Data

Data availability and accuracy is 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 a lack of data. Model complexity can give the impression of credibility but this is not always true.

It is often preferable to begin with simple models and then over time add additional complexity as justified based on the collection and analysis of additional data. This strategy makes

Fig. 10.3 Cause and effect diagram for estuary eutrophication due to excessive nutrient loadings (Borsuk et al. 2001)

efficient use of resources. It focuses efforts toward obtaining information and models that will reduce the uncertainty as the analysis proceeds. Models should be selected (simple vs. complex) in part based on 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 originate 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., each 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, which can vary over space and time within the area and time period being simulated, are typically not known with any precision. Chapter 12 reviews some of these loading models and their limitations.

10.4 Models of Water Quality Processes

Water quality models can be applied to many different types of water systems including streams, rivers, lakes, reservoirs, estuaries, coastal waters, and oceans. The models describe the main water quality processes and typically require the hydrologic 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 hydrologic 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 lake quality predictions is often more difficult than for streams, rivers, and estuaries. In coastal waters and oceans, larges scale flow patterns and tides are the most important transport mechanisms.

As with water quantity modeling, the development and application of water quality models is both a science as well as 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, and the time available for modeling and associated uncertainty and other analyses. The fact that most, if not all, water quality models cannot accurately predict what actually happens does not necessarily 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 water and land management policies or practices.

10.4.1 Mass Balance Principles

The basis principle of water quality models is that of mass balance. A water system can be divided into different segments or volumes elements (e.g., stream or river reaches, lake layers, estuary segments), also called "computational cells." For each segment or cell, there must be mass balance for each water quality constituent over time. Most water quality simulation models simulate quality over a consecutive series of discrete time period durations, Δt. Time is divided into discrete intervals 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 administrated. Components of the mass balance for a segment include: (1) changes by transport (Tr) into and out of the segment, (2) changes by physical or chemical processes (P) occurring within the segment and (3) changes by sources/discharges to or from the segment (S).

$$M_i^{t+\Delta t} = M_i^t + \Delta t \left(\frac{\Delta M_i}{\Delta t}\right)_{\text{Tr}} + \Delta t \left(\frac{\Delta M_i}{\Delta t}\right)_P + \Delta t \left(\frac{\Delta M_i}{\Delta t}\right)_S$$

$$(10.2)$$

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; $\left(\frac{\Delta M_i}{\Delta t}\right)_{\text{Tr}}$
- 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., waste loads, river discharges): $\left(\frac{\Delta M_i}{\Delta t}\right)_S$

Transport includes both advective and dispersive transport. Advective transport is the 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, includes the physical concept of molecular diffusion as it represents all transport that is not described by the advective transport.

Processes include physical processes such as reaeration 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.

Sources include the addition of mass by waste loads 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 modeled segment or volume element (the computational cell) is derived from monitoring data or a water quantity (possibly hydrodynamic) model.

To model the transport of substances over space, a water system can be divided into small segments or volume elements. The complete ensemble of all the segments or elements is called the "*grid*" or "s*chematization.*" Each computational cell is defined by its volume and its dimensions in one, two, or three directions (Δx, Δy, Δz) depending on the nature of the schematization (1D, 2D, or 3D). Note that the

values of Δx, Δy and Δz do not have to be equal. The computational cell can have any rectangular shape. A computational cell can share surface areas with other computational cells, the atmosphere and 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.

10.4.1.1   AdvectiveTransport

The advective transport, $T_{x_0}{}^A$ (M T^{-1}), of a constituent at a site x_0 is the product of the average water velocity, v_{x_0} (L T^{-1}), 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:

$$T_{x_0}{}^A = v_{x_0} \times A \times C_{x_0} \qquad (10.3)$$

10.4.1.2   Dispersive Transport

The dispersive transport, $T_{x_0}^D$ (M T^{-1}) across a surface area is assumed to be proportional to the concentration gradient $\frac{\partial C}{\partial x}\big|_{x=x_0}$ at site x_0 times the surface area A. Letting D_{x_0} (L^2 T^{-1}), be the dispersion or diffusion coefficient at site x_0:

$$T_{x_0}^D = -D_{x_0} \times A \times \frac{\partial C}{\partial x}\bigg|_{x=x_0} \qquad (10.4)$$

Dispersion is commonly based on Fick's diffusion law. The minus sign reflects the fact that dispersion causes net transport from higher to lower concentrations, 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

$$(\partial C/\partial x)|_x = \lim(\Delta x \to 0) \\ [(C_{x+0.5\Delta x} - C_{x-0.5\Delta x})/\Delta x] \qquad (10.5)$$

Dispersion coefficients should be calibrated or be obtained from calculations with turbulence models.

10.4.1.3   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, the one-dimensional equation results

$$M_i^{t+\Delta t} = M_i^t + \Delta t \times \left(v_{x_0} C_{x_0} - v_{x_0+\Delta x} C_{x_0+\Delta x} \right.$$
$$\left. -D_{x_0} \frac{\partial C}{\partial x}\bigg|_{x_0} + D_{x_0+\Delta x} \frac{\partial C}{\partial x}\bigg|_{x_0+\Delta x} \right) \times A$$

(10.6)

or equivalently:

$$M_i^{t+\Delta t} = M_i^t + \Delta t \times \left(Q_{x_0} C_{x_0} - Q_{x_0+\Delta x} C_{x_0+\Delta x} \right.$$
$$\left. -D_{x_0} A_{x_0} \frac{\partial C}{\partial x}\bigg|_{x_0} + D_{x_0+\Delta x} A_{x_0+\Delta x} \frac{\partial C}{\partial x}\bigg|_{x_0+\Delta x} \right),$$

(10.7)

where Q_{x_0} ($L^3 T^{-1}$) is the flow at site x_0.

If the previous equation is divided by the volume and the time interval Δt, then the following equation results in one dimension.

$$\frac{C_i^{t+\Delta t} - C_i^t}{\Delta t} = \frac{D_{x_{0+\Delta x}} \frac{\partial C}{\partial x}\big|_{x_0+\Delta x} - D_{x_0} \frac{\partial C}{\partial x}\big|_{x_0}}{\Delta x}$$
$$+ \frac{v_{x_0} C_{x_0} - v_{x_{0+\Delta x}} C_{x_{0+\Delta x}}}{\Delta x}$$

(10.8)

Taking the asymptotic limit $\Delta t \to 0$ and $\Delta x \to 0$, the advection–diffusion equation for one dimension results

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D \frac{\partial C}{\partial x} \right) - \frac{\partial}{\partial x}(vC) \qquad (10.9)$$

The finite volume computational method for transport can be used to solve the advection–diffusion equation. The accuracy of the method will be related to the size of Δx, A ($=\Delta y\,\Delta z$) and Δt.

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

$$\frac{\partial C}{\partial t} = D_x \frac{\partial^2 C}{\partial x^2} - v_x \frac{\partial C}{\partial x} + D_y \frac{\partial^2 C}{\partial y^2}$$
$$- v_y \frac{\partial C}{\partial y} + D_z \frac{\partial^2 C}{\partial z^2} - v_z \frac{\partial C}{\partial z} + S + f_R(C,t)$$

(10.10)

with dispersion coefficients D_j defined for each direction j. 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 "waste loads" (S): these source terms are additional inflows of water or mass. As many source terms as required may be added to Eq. 10.10. These could include small rivers, discharges of industries, sewage treatment plants, small waste load outfalls, etc.
- Reaction terms or "processes" (f_R).

Processes can be split into physical processes 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 Sect. 10.4.

Multiplying each term in Eq. 10.9 by the cross-sectional area A (L^2), the expression $DA(\partial C/\partial x) - vAC$ for a one-dimensional model, or its equivalent in Eq. 10.10 for a three-dimensional model, is termed the total flux ($M\,T^{-1}$). 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 concentrations zones. The coefficient of dispersion D ($L^2 T^{-1}$) 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 ($L T^{-1}$) across a cross-sectional area A.

The relative importance of dispersion and advection depends on how detailed 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 which assumes that 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 considered acceptable for natural aquatic systems.

10.4.2   Steady-State Models

Steady state means no change in the concentrations over time. In this case the left-hand side of Eq. 10.9 or 10.10, $\partial C/\partial t$, 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 Eq. 10.9 becomes

$$0 = D\partial^2 C/\partial x^2 - v\partial C/\partial x - kC \qquad (10.11)$$

Equation 10.11 can be integrated since river reach parameters $A, D, k, v,$ and Q are assumed constant. For a constant loading, W_C ($M T^{-1}$) at

site $x = 0$, the concentration C at any distance x will equal

$$C(x) = \begin{cases} (W_C/Qm)\exp[(v/2D)(1+m)x] & x \leq 0 \\ (W_C/Qm)\exp[(v/2D)(1-m)x] & x \geq 0 \end{cases}$$
$$(10.12)$$

where

$$m = \left(1 + \left(4kD/v^2\right)\right)^{1/2} \qquad (10.13)$$

Note from Eq. 10.13 that the parameter m is always equal or greater than 1. Hence the exponent of e in Eq. 10.12 is always negative. As the distance x increases in magnitude, either in the positive or negative direction, the concentration $C(x)$ will decrease. The maximum concentration C occurs at $x = 0$ and is W_C/Qm.

$$C(0) = W_C/Qm \qquad (10.14)$$

These equations are plotted in Fig. 10.4.

In flowing rivers not under the influence of tidal actions the dispersion is small. Assuming the dispersion coefficient D is 0, the parameter m defined by Eq. 10.13, is 1. When $D = 0$, the maximum concentration at $x = 0$ is W_C/Q.

$$C(0) = W_C/Q \quad \text{if} \quad D = 0. \qquad (10.15)$$
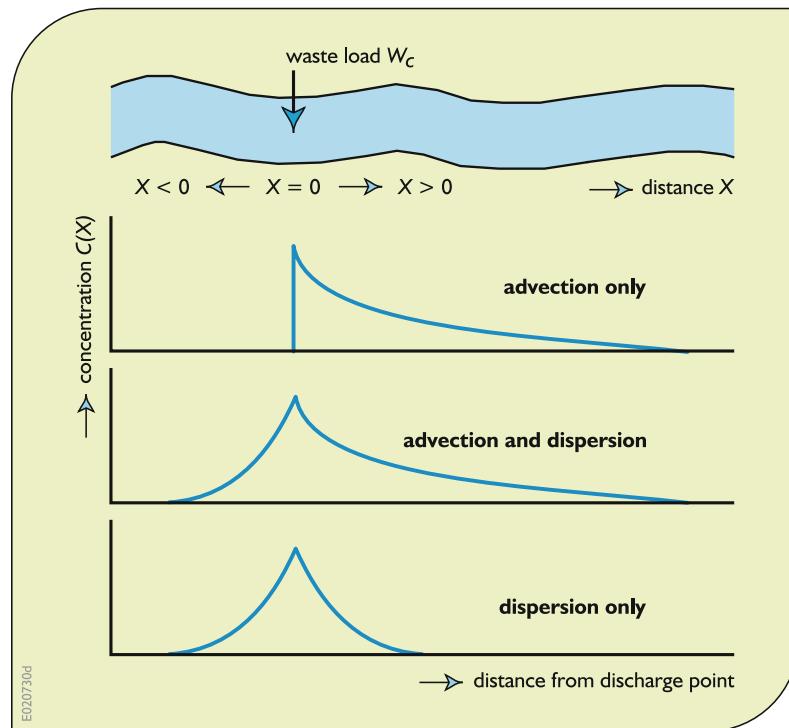
Assuming $D = 0$ and v, Q and $k > 0$, Eq. 10.12 becomes

$$C(x) = \begin{cases} 0 & x \leq 0 \\ (W_C/Q)\exp[-kx/v] & x \geq 0 \end{cases} \quad (10.16)$$

The above equation for $x > 0$ can be derived from Eqs. 10.12 and 10.13 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 Eq. 10.12 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 $x > 0$.

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,

Fig. 10.4 Constituent concentration distributions along a river or estuary resulting from a constant discharge of that constituent at a single point source in that river or estuary. The *top plot* represents Eq. 10.16, the *middle plot* represents Eq. 10.12, and the *bottom plot* represents Eq. 10.17

$Q = Av$, 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 Eq. 10.12 approaches $\pm x(k/D)^{1/2}$.

Hence for small velocities, Eq. 10.4 becomes

$$C(X) = \begin{cases} \left(W_C/2A(kD)^{1/2}\right)\exp\left[+x(k/D)^{1/2}\right] & x \le 0 \\ \left(W_C/2A(kD)^{1/2}\right)\exp\left[-x(k/D)^{1/2}\right] & x \ge 0 \end{cases}$$

$$(10.17)$$
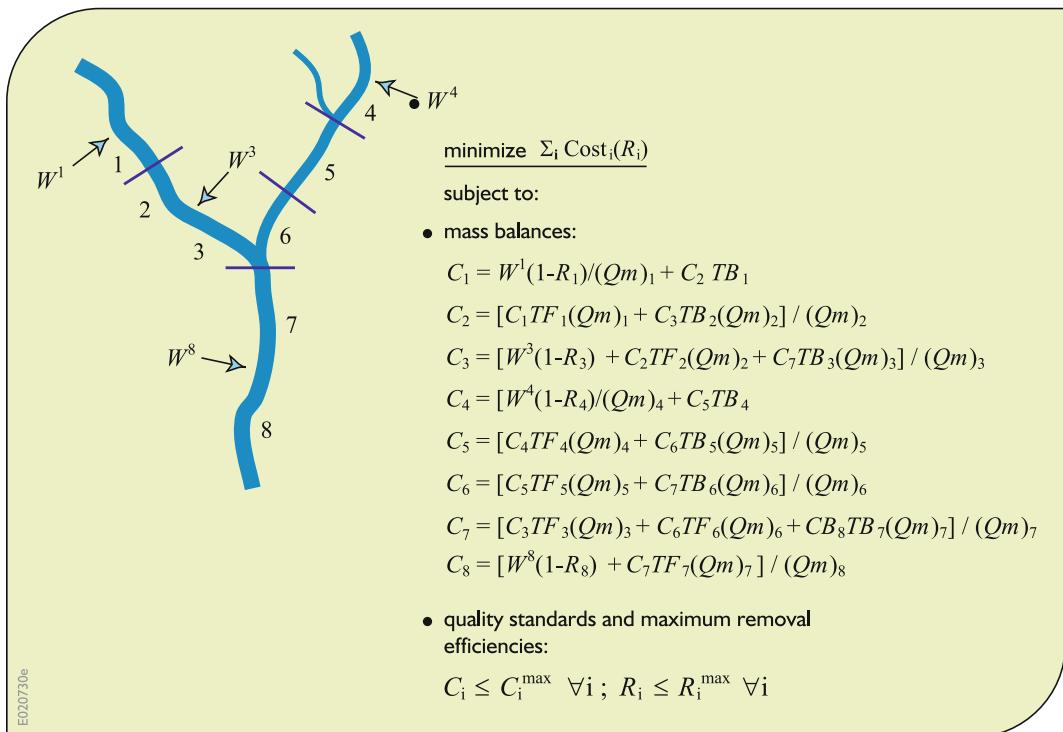
Here dispersion is much more important than advective transport and the concentration profile approaches a symmetric distribution, as shown in Fig. 10.4, about the point of discharge at $x = 0$.

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 10.5 shows a schematic of a river into which wastewater containing constituent C is being discharged at four sites. Assume maximum allowable concentrations of the constituent C are specified at each of those discharge sites. To estimate the needed 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 reach. These reaches are shown in Fig. 10.5.

In Fig. 10.5 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)

minimize $\sum_i \text{Cost}_i(R_i)$

subject to:

- mass balances:

$C_1 = W^1(1-R_1)/(Qm)_1 + C_2\,TB_1$

$C_2 = [C_1\,TF_1(Qm)_1 + C_3\,TB_2(Qm)_2]\,/\,(Qm)_2$

$C_3 = [W^3(1-R_3) + C_2\,TF_2(Qm)_2 + C_7\,TB_3(Qm)_3]\,/\,(Qm)_3$

$C_4 = [W^4(1-R_4)/(Qm)_4 + C_5\,TB_4$

$C_5 = [C_4\,TF_4(Qm)_4 + C_6\,TB_5(Qm)_5]\,/\,(Qm)_5$

$C_6 = [C_5\,TF_5(Qm)_5 + C_7\,TB_6(Qm)_6]\,/\,(Qm)_6$

$C_7 = [C_3\,TF_3(Qm)_3 + C_6\,TF_6(Qm)_6 + CB_8\,TB_7(Qm)_7]\,/\,(Qm)_7$

$C_8 = [W^8(1-R_8) + C_7\,TF_7(Qm)_7]\,/\,(Qm)_8$

- quality standards and maximum removal efficiencies:

$C_i \leq C_i^{\max}\ \ \forall i\,;\ \ R_i \leq R_i^{\max}\ \ \forall i$

Fig. 10.5 Optimization model for finding constituent removal efficiencies, R_i, at each discharge site i that result in meeting stream quality standards, $C_i^{\max}$, at least total cost

transfer coefficient, TF_i, equals the applicable part of Eq. 10.12,

$$\text{TF}_i = \exp[(v/2D)(1-m)x] \qquad (10.18)$$

as does the upstream (backward) transfer coefficient, TB_i.

$$\text{TB}_i = \exp[(v/2D)(1+m)x] \qquad (10.19)$$

The parameter m is defined by Eq. 10.13.

Finding the cost solution of a model such as shown in Fig. 10.5 does not mean that the least-cost wasteload allocation plan will be implemented, but such information can help 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 program. Non economic objectives can also be used to obtain other tradeoffs.

The model in Fig. 10.5 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.

10.4.3 Design Streamflows for Setting and Evaluating Quality Standards

In streams and rivers, the water quality may vary significantly depending on the stream or river flow and its quality prior to wastewater discharges. If waste load discharges are fairly constant, a high flow of high quality serves to dilute the waste concentration while contaminant concentrations of low flows may become undesirably high. It is therefore common practice to pick
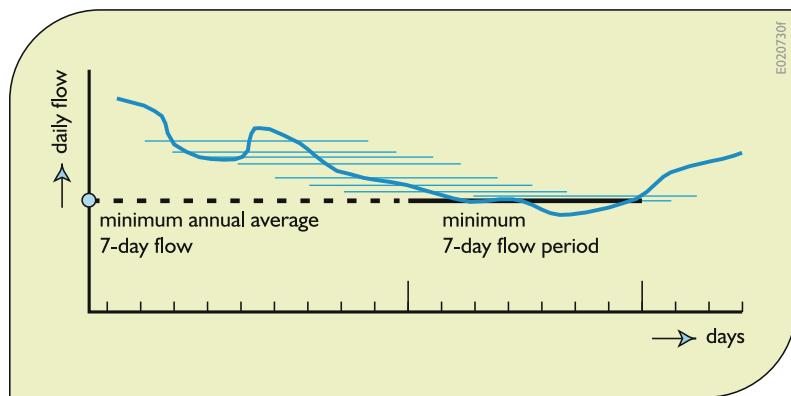
a more critical low-flow condition for judging whether or not ambient water quality standards are being met. This can also be seen from Eqs. 10.12, 10.14, 10.15, and 10.16. This often is the basis for the assumption that the smaller (or more critical) the design flow, the more likely it is that the stream quality standards will be met in practice. 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, that is used in parts of North America, is the minimum 7-day average flow expected to be lower only once in 10 years on average. Each year the lowest 7-day average flow is determined, as shown in Fig. 10.6. The sum of each of the 365 sequences of seven average daily flows is divided by 7 and the minimum value is selected. This is the minimum annual average 7-day flow for the year.
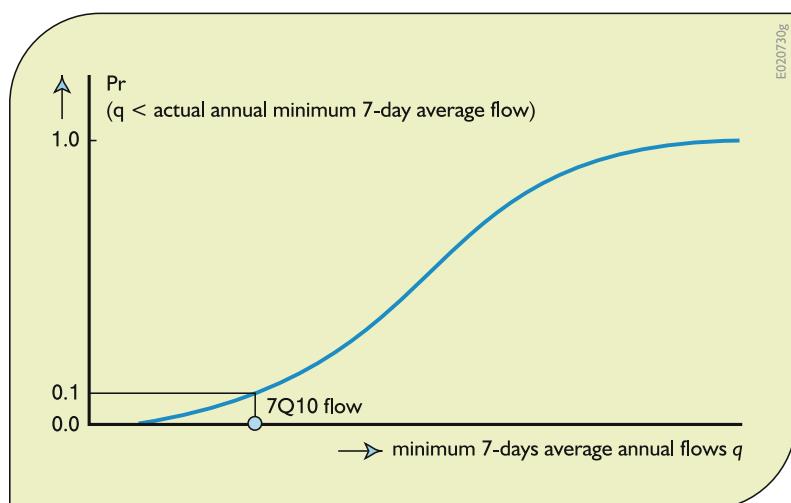
These minimum 7-day average flows for each year of record define a probability distribution, whose cumulative probabilities can be plotted. As illustrated in Fig. 10.7, the particular flow on the cumulative distribution that has a 90% chance of being exceeded is the design flow. It is the minimum annual average 7-day flow expected once in 10 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 about a one percent chance that any daily flow will be less than this 7Q10 flow.

Fig. 10.6 Portion of annual flow time series showing low flows and the calculation of average 7 and 14-day flows

Fig. 10.7 Determining the minimum 7-day annual average flow expected once in 10 years, designated 7Q10, from the cumulative probability distribution of annual minimum 7-day average flows

Consider now any one of the river reaches shown in Fig. 10.5. Assume an initial loading of constituent mass, $M/\Delta t$, exists at the beginning of the reach. As the reach flow, Q, increases and the mass loading stays the same, the initial concentration, M/Q, will decrease. However, the flow velocity will increase, and thus the time, Δt, 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. Hence it is possible that ambient water quality standards that are met during low flow conditions may not be met under higher flow conditions, conditions that are observed much more frequently. Figure 10.8 illustrates how this might happen. This does not suggest that low flows should not be considered when allocating waste loads, 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 they are during less frequent low-flow conditions.

Figure 10.8 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 higher flow conditions.
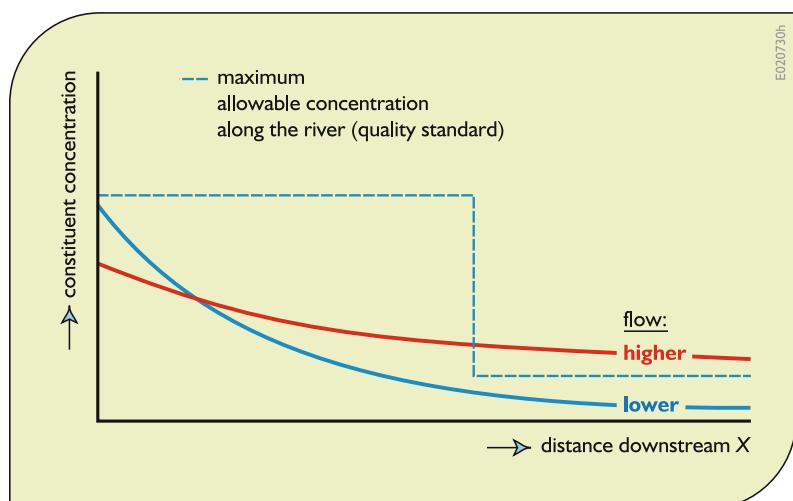
10.4.4   Temperature

Temperature impacts almost all water quality processes taking place in water bodies. For this reason modeling temperature may be important when the temperature varies 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 shortwave solar radiation, longwave 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 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 ρ (g/cm^3), the heat capacity of water, c_p (cal/g/°C), and the water depth h (cm).

Fig. 10.8 Increasing streamflows decreases initial concentrations but may increase downstream concentrations

Assuming the net heat input, $K_H(T_e - T)$ (cal/cm²/day), is proportional to the difference of the actual temperature, T, and the equilibrium temperature, T_e (°C),

$$dT/dt = K_H(T_e - T)/\rho c_p h \qquad (10.20)$$

The overall heat exchange coefficient, K_H (cal/cm²/day/°C), is determined in units of W/m²/°C (1 cal/cm²/day °C = 0.4840 W/m²/°C) from empirical relationships that include wind velocity U_w (m/s), dew point temperature T_d (°C), 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 shortwave solar radiation H_s (cal/cm²/day),

$$T_e = T_d + (H_s/K_H) \qquad (10.21)$$

This model simplifies the mathematical relationships of a complete heat balance and requires less data.

10.4.5 Sources and Sinks

Sources and sinks of water quality constituents include the physical and biochemical processes that are represented by the term S in Eq. 10.10. External inputs of each constituent would have the form W/Q, where W (M T^{-1}) is the loading rate of the constituent and Q represents the flow of water into which the mass of waste W is discharged.

10.4.6 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 magnitude of concentration C_i, of constituent i, then

$$dC_j/dt = a_{ij}k_i\theta_i^{(T-20)}C_i, \qquad (10.22)$$

where θ_i is 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 Eq. 10.22. 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, i.e., dilution.

The typical values for the rate coefficients k_c and temperature coefficients θ_i of some constituents C are in Table 10.1. 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 $k_B = 2.3/t_{90}$.

10.4.7 Dissolved Oxygen

DO concentration is a common indicator of the health of the aquatic ecosystem. DO was originally modeled by Streeter and Phelps (1925). Since then a number of modifications and extensions of the model have been made depending on the number of sinks and sources of DO being considered and how processes involving the nitrogen cycle and phytoplankton, are being modeled, as illustrated in Fig. 10.9.

The sources of DO in a water body include reaeration from the atmosphere, photosynthetic oxygen production from aquatic plants, denitrification, and DO inputs. The sinks include oxidation of carbonaceous and nitrogenous material, sediment oxygen demand and respiration by aquatic plants.

Table 10.1 Typical values of the first-order decay rate, k, and the temperature correction factor, θ, for some constituents

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)

$$\Delta O_2/\Delta t = \text{loads} + \text{transport} + \text{reaeration}$$
$$+ \text{net primary production} + \text{denitrification}$$
$$- \text{mineralization} - \text{nitrification} - \text{SOD}$$

The rate of reaeration is assumed to be proportional to the difference between the saturation concentration, DO_{sat} (mg/l), and the concentration of DO, DO (mg/l). The proportionality coefficient is the reaeration 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, θ, 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 reaeration rate constant is a sensitive parameter. There have been numerous equations developed to define this rate constant. Table 10.2 lists some of them.
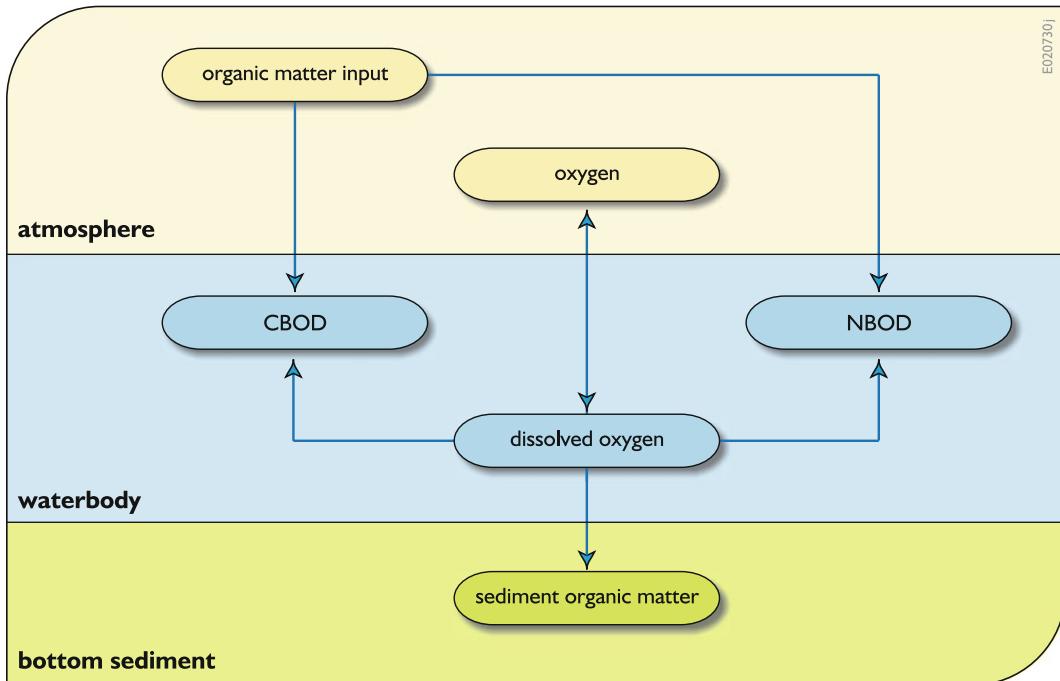
The saturation concentration, DO_{sat}, of oxygen in water is a function of the water temperature and salinity [chloride concentration, Cl (g/m^3)], and can be approximated by

$$DO_{sat} = \left\{ 14.652 - 0.41022T + (0.089392T)^2 \right.$$

$$\left. - (0.042685T)^3 \right\} \left\{ 1 - (Cl/100000) \right\} \quad (10.23a)$$

Table 10.2   Some equations for defining the reaeration rate constant, k_r (1/day)

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)

E020813b
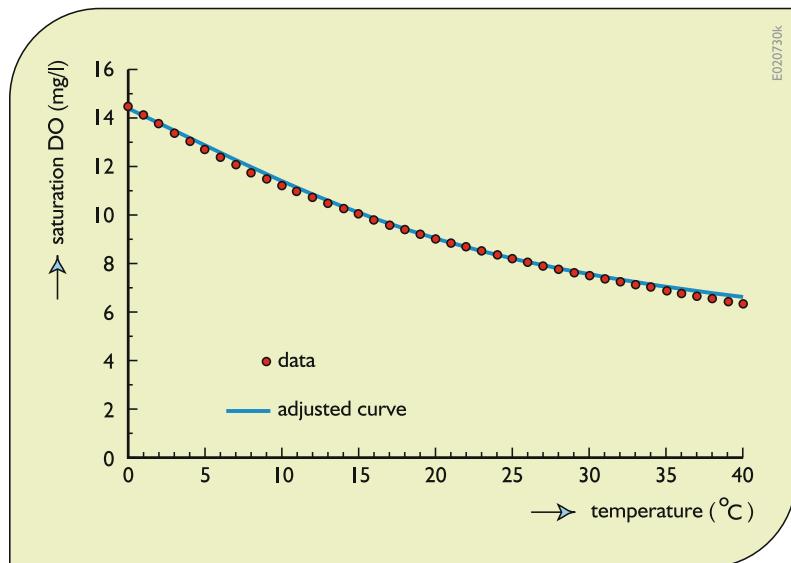
Fig. 10.9 The dissolved oxygen interactions in a water body, showing the decay (satisfaction) of carbonaceous, nitrogenous and sediment oxygen demands and water body reaeration or deaeration (if supersaturated occurs at the air–water interface)

Fig. 10.10 Fitted curve to the saturation dissolved oxygen concentration (mg/l) as a function of temperature (°C)

Elmore and Hayes (1960) derived an analytical expression for the DO saturation concentration, DO_{sat} (mg/l), as a function of temperature (T, °C):

$$DO_{sat} = 14.652 - 0.41022T + 0.007991T^2 - 0.000077774T^3$$

$$(10.23b)$$

Fitting a second-order polynomial curve to the data presented in Chapra (1997) results in

$$DO_{sat} = 14.407 - 0.3369T + 0.0035T^2$$

$$(10.23c)$$

as is shown in Fig. 10.10.

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.

The biochemical oxygen demand results from carbonaceous organic matter (CBOD, mg/l) and 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 modeled 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 θ_{CBOD} and θ_{NBOD}, respectively.

The sediment oxygen demand (SOD) (mg/liter/day) is usually expressed as a zero-order reaction, i.e., a constant demand. One important feature in modeling 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 DO model with CBOD, NBOD, and SOD is

$$dDO/dt = -k_{CBOD}\theta_{CBOD}^{(T-20)}CBOD - k_{NBOD}\theta_{NBOD}^{(T-20)}NBOD + k_r q_r^{(T-20)}(DO_{sat} - DO) - SOD$$

$$(10.24)$$

$$dCBOD/dt = -k_{CBOD}\theta_{CBOD}^{(T-20)}CBOD \quad (10.25)$$

$$dNBOD/dt = -k_{NBOD}\theta_{NBOD}^{(T-20)}NBOD \quad (10.26)$$

The mean and range values for coefficients included in these DO models are in Table 10.3.

Table 10.3 Typical values of parameters used in the dissolved oxygen models

parameter	value		units
k_r, slow, deep rivers	0.1-0.4	- a	l/day
k_r, typical conditions	0.4-1.5	- a	l/day
k_r, swift, deep rivers	1.5-4.0	- a	l/day
k_r, swift, shallow rivers	4.0-10.0	- a	l/day
k_{CBOD}, untreated discharges	0.35 (0.20-0.50)	- b	l/day
k_{CBOD}, primary treatment	0.20 (0.10-0.30)	- b	l/day
k_{CBOD}, activated sludge	0.075 (0.05-0.10)	- b	l/day
θ_{CBOD}	1.04	- a	—
	1.047	- a	—
	1.04 (1.02-1.09)	- c	—
θ_r	1.024 (1.005-1.030)	- c	---

sediment oxygen demand [*]	value		units
municipal sludge (outfall vicinity)	4 (2-10)	- c d	$gO_2/m^2/day$
municipal sewage sludge	1.5 (1-2)	- c d	$gO_2/m^2/day$
sandy bottom	0.5 (0.2-1.0)	- c d	$gO_2/m^2/day$
mineral soils	0.07 (0.05-0.1)	- c d	$gO_2/m^2/day$
natural to low pollution	0.1-10.0	- a	$gO_2/m^2/day$
moderate to heavy pollution	5-10	- a	$gO_2/m^2/day$

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)

10.4.8   Nutrients and Eutrophication

Eutrophication is the progressive process of nutrient enrichment of water systems. An increase in nutrients leads to an increase in the productivity of the water system that 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
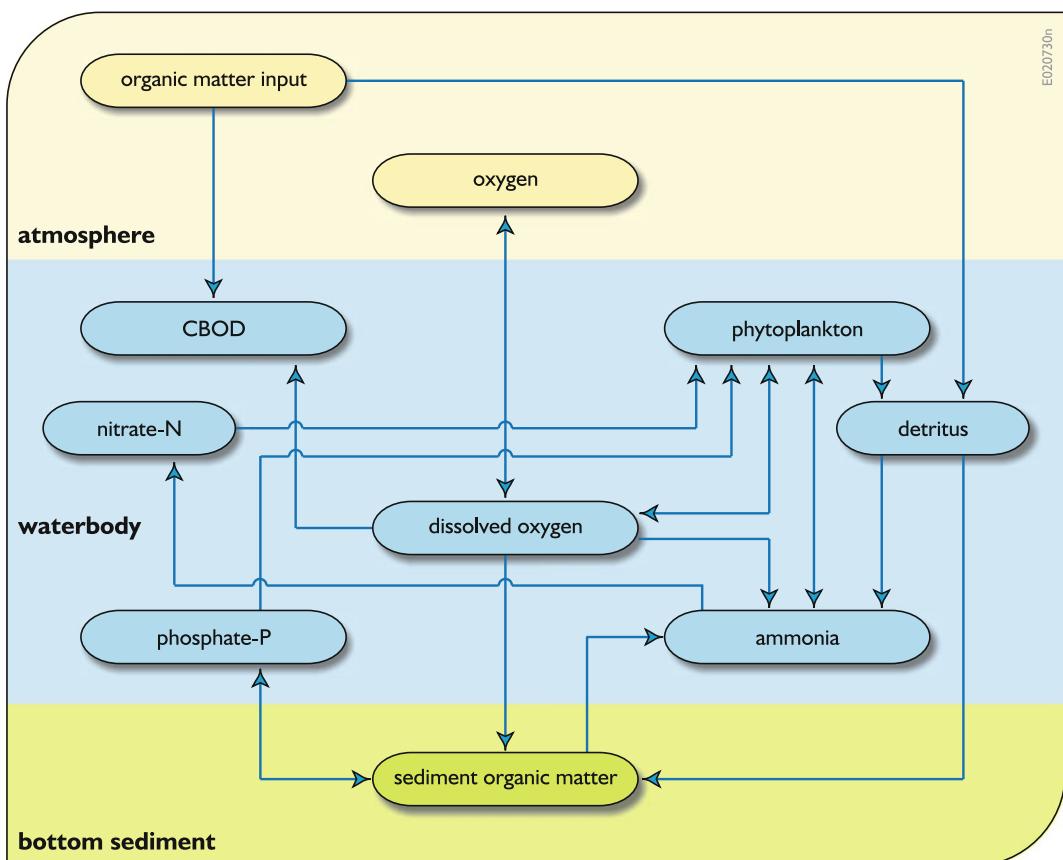
water it is called an algae bloom. Excessive algal biomass could affect the water quality, especially if it causes anaerobic conditions and thus impairs the drinking, recreational, and ecological uses.

The eutrophication component of the model relates the concentration of nutrients and the algal biomass. For example, as shown in Fig. 10.11, consider the growth of algae A (mg/l—not to be confused with area A used in previous equations), 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.
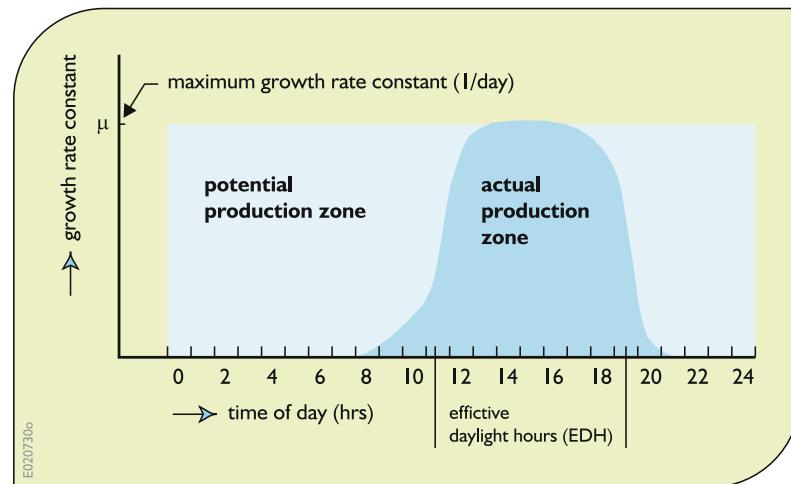
To account for this, algal growth is commonly modeled as a Michaelis–Menten multiplicative effect, i.e., the nutrients have a synergistic effect. Model parameters include a maximum algal growth rate μ (1/day) times the fraction of a day, f_d, that rate applies (Fig. 10.12), the half saturation constants K_P and K_N (mg/l) (shown as K_c in Fig. 10.13) 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 be in proportion to their contents in the algae
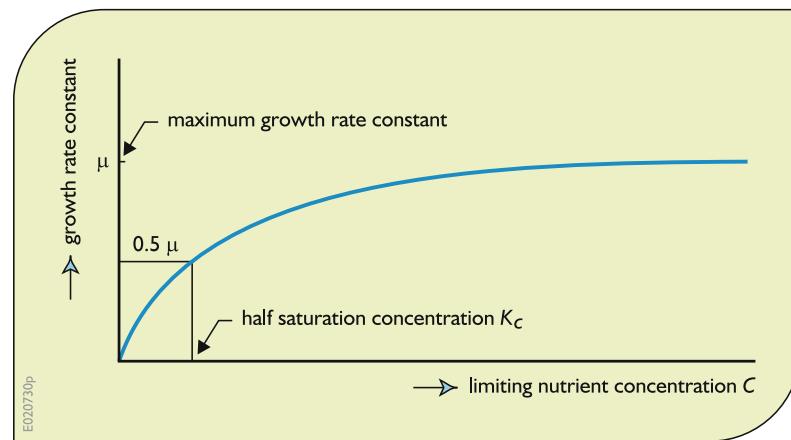
Fig. 10.11 The dissolved oxygen, nitrogen, and phosphorus cycles, and phytoplankton interactions in a water body, showing the decay (satisfaction) of carbonaceous and sediment oxygen demands, reaeration or deaeration of oxygen at the air–water interface, ammonification of organic nitrogen in the detritus, 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

Fig. 10.12 Calculation of the fraction, f_d, of the maximum growth rate constant, μ, to use in the algal growth equations. The fraction f_d is the ratio of actual production zone/potential production zone

Fig. 10.13 Defining the half saturation constant for a Michaelis–Menten model of algae. The actual growth rate constant $= \mu C/(C + K_C)$

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 Fig. 10.11 can be written as follows:

For algae biomass

$$dA/dt = \mu f_d \theta_A^{(T-20)} [P/(P + K_P)][N_n/(N_n + K_N)]A - e\theta_A^{(T-20)}A$$

$$(10.27)$$

For organic phosphorus

$$dP_o/dt = -k_{op}\theta_{op}^{(T-20)}P_o \qquad (10.28)$$

For phosphate phosphorus

$$dP/dt = -\mu f_d\theta_A^{(T-20)}[P/(P+K_P)][N_n/(N_n+K_N)]a_PA \qquad (10.29)$$

For organic nitrogen

$$dN_o/dt = -k_{on}\theta_{on}^{(T-20)}N_o \qquad (10.30)$$

For ammonia–nitrogen

$$dN_a/dt = -\mu f_d\theta_A^{(T-20)}[P/(P+K_P)][N_n/(N_n+K_N)]a_AA$$
$$+ k_{on}\theta_{on}^{(T-20)}N_o - k_a\theta_a^{(T-20)}N_a \qquad (10.31)$$

For nitrate–nitrogen

$$dN_N/dt = -\mu f_d\theta_A^{(T-20)}[P/(P+K_P)][N_n/(N_n+K_N)]a_NA$$
$$+ k_a\theta_a^{(T-20)}N_a - k_n\theta_A^{(T-20)}N_n \qquad (10.32)$$

For DO

$$dDO/dt = -k_{CBOD}\theta_{CBOD}^{(T-20)}CBOD - 4.57k_a\theta_a^{(T-20)}N_a$$
$$- 2k_{op}\theta_{op}^{(T-20)}P_o + (1.5\,\mu f_d - 2e)\theta_A^{(T-20)}A$$
$$+ k_r\theta_r^{(T-20)}(DO_{sat} - DO) - SOD \qquad (10.33)$$

Representative values of the coefficients for this model are in Table 10.4.

Because of the growth of phytoplankton cannot occur without nutrients, the eutrophication modeling must be coupled with that of nutrients. Nutrient modeling 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

Table 10.4 Typical values of coefficients in the eutrophication model

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 stoichiometric ratio	0.012-0.015	- c	mg P/mg A
a_N stoichiometric ratio	0.08-0.09	- c	mg NO_3/mg A
μ maximum algae growth rate	1.5 (1.0-2.0)	- b	l/day
e death algae rate	0.2-8	- c	l/day
	0.1 (0.05-0.025)	- b	l/day
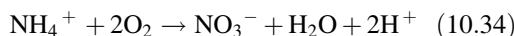
a - Thomann and Mueller (1987)

b - Schnoor (1996)

c - Bowie et al. (1985)

NO$_3$–N) as the main forms of dissolved nitrogen in water. Furthermore, nitrogen is present in algae, as well as in detritus, resulting from algae mortality, and in 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. Nitrification is the conversion of ammonium to nitrite and finally nitrate, requiring the presence of oxygen

$$NH_4{}^+ + 2O_2 \rightarrow NO_3{}^- + H_2O + 2H^+ \quad (10.34)$$

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 systems

$$NO_3{}^- \rightarrow N_2 \quad (10.35)$$

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). Similar to 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.

10.4.9 Toxic Chemicals
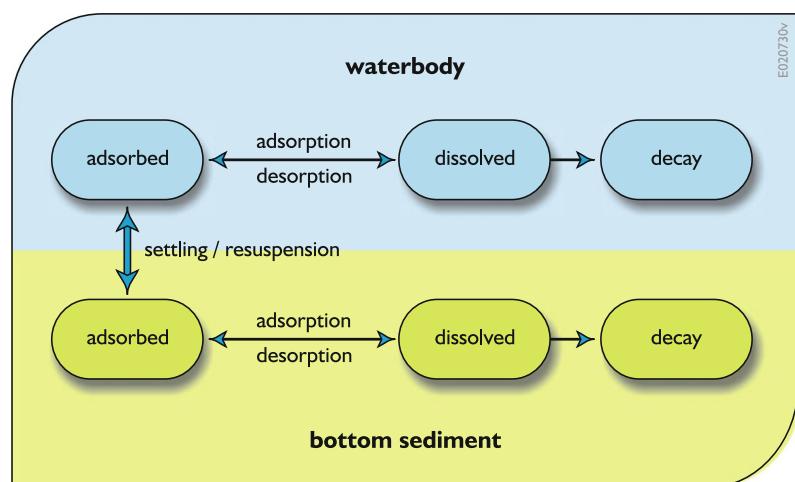
Toxic chemicals, also referred to as "micropollutants," are substances that at low concentrations can impair the reproduction and growth of organisms including fish and human beings. These substances include heavy metals, many synthetic organic compounds (organic micropollutants), and radioactive substances.
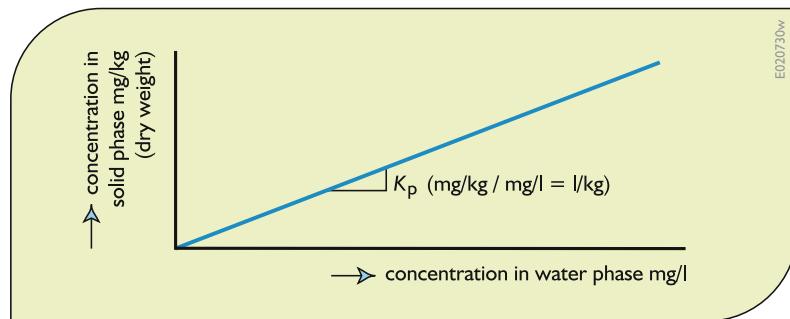
10.4.9.1 Adsorbed and Dissolved Pollutants

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 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 10.14 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

Fig. 10.14 Schematic of the adsorption/desorption and decay processes of various toxic chemicals in water bodies and bottom sediments

Fig. 10.15 Defining the partition coefficient K_p (liters/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

phases in both the water body and in the bottom sediments.

The adsorption–desorption model assumes (conveniently but not always precisely) that an equilibrium exists between the dissolved (in water) and absorbed (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 (l/kg). This is shown in Fig. 10.15.

Each partition coefficient K_p (liters per kilogram or l/kg) is defined as the ratio of the particulate concentration C_p' of a micropollutant (mg/kg C) divided by the dissolved concentration C_d' of a micropollutant (mg/l water).

$$K_p = C_p'/C_d' \qquad (10.36)$$

Representative values of partition coefficients K_p are in Table 10.5.

The presence of a micropollutant in water 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 micropollutants is largely determined by adsorption to particulate matter, suspended inorganic and organic matter (including phytoplankton) have 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 modeled part of the water system by burial.

The rates of settling and resuspension of micropollutants are proportional to the rates for particulate matter. An additional process called bioturbation leads to redistribution of the micropollutant among sediment layers. Bioturbation is caused by physical activity of organisms, and affects both the particulate and dissolved phases but with different rates. Bioturbation is taken into account by means of dispersion coefficients.

For modeling 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 absorbed fraction (f_{aw}).

$$C_w = (f_{dw} + f_{aw})C_w \qquad (10.37)$$

In turn, the adsorbed fraction of a micropollutant is composed of the fractions adsorbed to inorganic particulate matter, f_{im}, dead particulate organic matter, f_{poc}, and algae, f_{alg}. The total

Table 10.5 Typical values of partition coefficients in toxic chemical model from Thomann and Mueller (1987)

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

micropollutant concentration, C_w (mg/m³) is the sum of all these fractions.

$$C_w = (f_{dw} + f_{im}, + f_{poc} + f_{alg})C_w \qquad (10.38)$$

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

$$DC_w = f_{dw}C_w \qquad (10.39)$$

$$AC_w = f_{aw}C_w, \qquad (10.40)$$

where the fractions

$$f_{dw} = 1/(1 + K_p\text{SS}) \qquad (10.41)$$

$$f_{aw} = 1 - f_{dw} = K_p\text{SS}/(1 + K_p\text{SS}) \qquad (10.42)$$

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).

$$DC_s = f_{ds}C_s \qquad (10.43)$$

$$AC_s = f_{as}C_s \qquad (10.44)$$

These fractions are dependent on the sediment porosity, ϕ, and density, ρ_s (kg/l).

$$f_{ds} = 1/[\phi + \rho_s(1 - \phi)K_p] \qquad (10.45)$$

$$f_{as} = 1 - f_{ds}$$
$$= ([\phi + \rho_s(1 - \phi)K_p]-1)/[\phi + \rho_s(1 - \phi)K_p] \qquad (10.46)$$

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

$$dC_w/dt = -k_w\theta_w^{(T-20)}f_{dw}C_w - f_{aw}C_wS + f_{as}C_sr \qquad (10.47)$$

$$dC_s/dt = -k_s\theta_s^{(T-20)}f_{ds}C_s + f_{aw}C_wS - f_{as}C_sr \qquad (10.48)$$

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 θ 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 θ. Suspended solids settling and resuspension can be determined at each day from a sediment model.

10.4.9.2 Heavy Metals

The behavior of heavy metals in the environment depends on their inherent chemical properties. Heavy metals can be divided into different categories depending on their dissolved form and redox (reduction oxidation) 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 sulfhides under chemically reducing conditions. These sulfhides generally settle in bottom sediments and are essentially ecologically unavailable. 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 Eqs. 10.39–10.42. 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 modeled 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 sulfhides or hydroxides can be modeled. 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.

10.4.9.3 Organic Mircropollutants

Organic micropollutants generally are biocides (such as pesticides or herbicides), solvents or combustion products and include substances such as hexachlorohexane, hexachlorobenzene, PCB's or polychlorobiphenyls, benzo-a-pyrene and fluoranthene (PAH's or polycyclic aromatic hydrocarbons), diuron and linuron, atrazine and simazine, mevinfos and dichlorvos, and dinoseb.

The short-term fate of organic micropollutants 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 micropollutant 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 micropollutants are generally poorly soluble in water and prefer to absorb to particulate matter in the water, especially particulate organic matter and algae. Therefore, the fractions of a micropollutant adsorbed to inorganic matter, f_{im}, dead particulate organic matter, f_{poc}, the dissolved fraction of a micropollutant, f_d, and algae, f_{alg}, add up to the total micropollutant concentration, C (mg/m^3).

$$C = (f_d + f_{im} + f_{poc} + f_{alg})C \qquad (10.49)$$

The fractions are functions of the partition coefficients K_p [for algae (m^3/g C), for inorganic matter (m^3 g DW^{-1}) and for dead particulate organic matter (m^3/g C)], the individual concentrations C [for algae biomass (g C/m^3), for dissolved (in water) inorganic matter (g/m^3) and for dead particulate organic matter (g C/m^3)], and the porosity ϕ (m^3 water/m^3 bulk). In surface water the value for porosity is 1.

$$f_d = \phi/[\phi + K_{palg}C_{alg} + K_{pim}C_{im} + K_{ppoc}C_{poc}] \qquad (10.50)$$

$$f_{im} = (1 - f_d)K_{pim}C_{im}/[K_{palg}C_{alg} + K_{pim}C_{im} + K_{ppoc}C_{poc}] \qquad (10.51)$$

$$f_{poc} = (1 - f_d)K_{ppoc}C_{poc}/[K_{palg}C_{alg} + K_{pim}C_{im} + K_{ppoc}C_{poc}] \qquad (10.52)$$

$$f_{alg} = (1 - f_d - f_{im} - f_{poc}) \qquad (10.53)$$

In terms of bulk measures, each partition coefficient K_p (see Eq. 10.36) also equals the porosity ϕ times the bulk particulate concentration C_p (mg/m^3 bulk) divided by the product of the dissolved (mg/l bulk) and particulate (mg/m^3 bulk) bulk concentrations, C_d C_s, all times 10^6 mg/kg.

$$K_p = 10^6 \phi \, C_p/(C_d \, C_s) \qquad (10.54)$$

Partitioning can be simulated based on the above equilibrium approach or according to slow sorption kinetics. For the latter, the rate, dC_p/dt, of adsorption or desorption (mg/m^3/day) depends on a first-order kinetic constant k_{sorp} (1/day) for adsorption and desorption times the difference between equilibrium particulate concentration C_{pe} of a micropollutant (mg/m^3 bulk) and the actual particulate concentration C_p (mg/m^3 bulk) of a micropollutant.

$$dC_p/dt = k_{sorp}(C_{pe} - C_p) \qquad (10.55)$$

The kinetic constant for sorption is not temperature dependent. All other kinetic constants for micropollutants are temperature dependent.

Mass balance equations are similar for all micropollutants except for the loss processes.

Metals are conservative substances. They can be transformed into various species either through complexation, adsorption, or precipitation. Organic micropollutants are lost by volatilization, biodegradation, photolysis, hydrolysis, and overall degradation. Most of these processes are usually modeled as first-order processes, with associated rate constants.

The volatization rate, dC_d/dt (mg/m^3/day) of dissolved micropollutant concentrations, C_d (mg/m^3 water) in water depends on an overall transfer coefficient, k_{vol} (m/day), for volatilization and the depth of the water column, H (m).

$$dC_d/dt = -k_{vol}C_d/H \qquad (10.56)$$

The numerator ($k_{vol} C_d$) 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 modeled as

$$dC/dt = -kC, \qquad (10.57)$$

where C is the total concentration of a micropollutant (mg/m^3), and k is a (pseudo) temperature dependent first-order kinetic rate constant for biodegradation, photolysis, hydrolysis or overall degradation (1/day). This is similar to Eq. 10.22.

10.4.9.4   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, the 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 modeled as a first-order process involving a kinetic radioactive decay constant, k_{dec} (1/day). 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

$$Ac = 10^{-3} N_A k_{dec} C_R / [86400 \text{ Mw}], \quad (10.58)$$

where Ac = activity of the radionuclide (Bq/m^3/s)
N_A = Avogadro's number (6.02 × 10^{23} mol)
Mw = molecular weight of the radionuclide (g/mole)

10.4.10   Sediments

Sediments in water 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 (2–20 mm) or sand (0.07–2 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 the 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 behavior of this fine-grained suspended matter impacts water quality. First, turbidity and its effect on the underwater light is an important environmental condition for algae growth. The presence of suspended sediment increases the attenuation of light in the water column that 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.

10.4.11   Processes in Lakes and Reservoirs

The water quality modeling principles discussed above are applicable to different types of water systems such as streams, rivers, lakes, estuaries, and even coastal or ocean waters. This section presents some of the unique aspects of water quality modeling in lakes. The physical character

and water quality of rivers draining into lakes and reservoirs are governed in part by the velocity and the 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. Portions of the sediment and other material 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. Greater opportunities for the growth of algae (phytoplankton) and the development of eutrophication are present.

Reservoirs typically receive larger inputs of water, as well as soil and other materials carried in rivers than lakes. As a result, reservoirs may receive larger pollutant loads than lakes. 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 than do lakes. 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

- 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), and
- acidity (a neutral pH of 7 is preferred).

Many lakes and reservoirs receive discharges of waste chemical compounds from industry, some with toxic or deleterious effects on humans and/or other water-dependent organisms and products. Some of these pollutants 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.
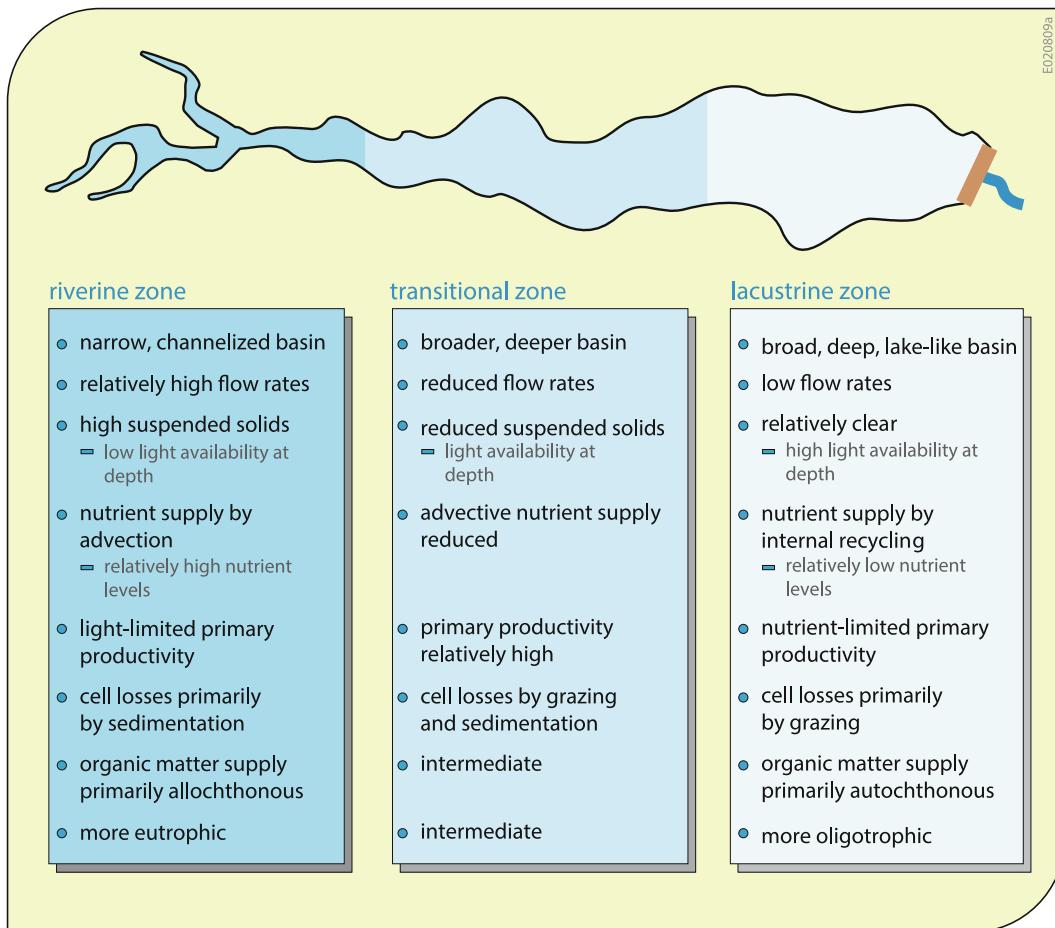
There can be major differences between deep and shallow lakes or reservoirs. Deep lakes, particularly in nontropical regions, usually have poorer water quality in lower layers, due to stratification (see Sect. 4.11.3). Shallow lakes do not exhibit this depth differentiation in quality. Their more shallow, shoreline areas have relatively poorer water quality because those sites are where pollutants are discharged and have a greater potential for disturbance of bottom muds, etc. 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 Fig. 10.16.

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 zones (Fig. 10.16). The relative size and volume of the three zones can vary greatly in a given reservoir.

10.4.11.1 Water Quality Changes and Impacts

Dams can produce changes in the downstream river channels below them. These are quite unlike downstream changes from lakes. Because reservoirs act as sediment and nutrient traps, the water at the downstream dam end of a reservoir is typically of higher quality than water entering the reservoir.

E020809a

riverine zone

- narrow, channelized basin
- relatively high flow rates
- high suspended solids
  - low light availability at depth
- nutrient supply by advection
  - relatively high nutrient levels
- light-limited primary productivity
- cell losses primarily by sedimentation
- organic matter supply primarily allochthonous
- more eutrophic

transitional zone

- broader, deeper basin
- reduced flow rates
- reduced suspended solids
  - light availability at depth
- advective nutrient supply reduced
- primary productivity relatively high
- cell losses by grazing and sedimentation
- intermediate
- intermediate

lacustrine zone

- broad, deep, lake-like basin
- low flow rates
- relatively clear
  - high light availability at depth
- nutrient supply by internal recycling
  - relatively low nutrient levels
- nutrient-limited primary productivity
- cell losses primarily by grazing
- organic matter supply primarily autochthonous
- more oligotrophic

Fig. 10.16 Longitudinal zonation of water quality and other variables in reservoirs

This higher quality water subsequently flows into the downstream river channel below the dam. This 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 can be picked up and transported as it moves downstream. Because it contains less sediment, the discharged "hungry" water can scour and erode the streambed and banks, picking up new sediment as it continues downstream. This scouring effect can negatively impact 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 water supplies, have water release or discharge structures located at different vertical levels in their dams (Fig. 10.17). This allows for the withdrawal or discharge of water from different layers within the reservoir, so-called "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 can impact both flora and fauna and possibly even the geomorphology of the stream or river.

Fig. 10.17 A multiple outlet reservoir can be better used to regulate the temperature and water quality downstream and possibly the sediment in the reservoir

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.

Agricultural and urban runoff 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) adversely impact lake ecosystems. When temperature and light conditions are favorable, excessive nutrients stimulate population explosions of undesirable algae and aquatic weeds. After they die the algae sink to the lake bottom, where bacteria consume DO as they decompose the algae. Fish kills and foul odors may result if dissolved is depleted.

Heavy metals are another major cause of lake quality impairment. Heavy metals accumulate in fish tissue. Since it is difficult to measure heavy metals (e.g., mercury) in ambient water and since they accumulate in fish tissue, fish samples are commonly used to monitor heavy metal contamination. Common sources of heavy metals are "smoke-stack" industries, including power plants, whose airborne discharges of mercury eventually end up in our water supplies.

In addition to nutrients and metals siltation, enrichment by organic wastes that deplete oxygen and noxious aquatic plants impact lakes and reservoirs. Often, several pollutants and processes impact 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 hydrologic modifications are also sources of lake pollutants.

10.4.11.2   Lake Quality Models

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.

Perhaps the simplest way to begin modeling 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 (1/day) into a lake having a volume V (m³) and inflow and outflow rate of Q (m³/day), then the rate of change in the concentration C (g/m³/day) is

$$dC/dt = (1/V)(W_C - QC - k_C CV) \quad (10.59)$$

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 that previous time period, $t-1$. For a period duration of Δt days,

$$\begin{aligned} C(t) = &[W_C/(Q + k_C V)][1 - \exp\{-\Delta t((Q/V) + k_C)\}] \\ &+ C(t-1)\exp\{-\Delta t((Q/V) + k_C)\} \end{aligned}$$
$$(10.60)$$

The equilibrium concentration, C_e, can be obtained by setting the rate, dC/dt, in Eq. 10.59 to 0. The net result is

$$C_e = W_C/(Q + k_C V) \quad (10.61)$$

The time, t_α, since the introduction of a mass input W_C that is required to reach a given fraction α of the equilibrium concentration (i.e., $C(t)/C_e = \alpha$) is

$$t_\alpha = -V[\ln(1 - \alpha)]/(Q + K_C V) \quad (10.62)$$

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

$$dC/dt = -C(Q + k_C V)/V \quad (10.63)$$

Integrating this equation, the concentration $C(t)$ is

$$C(t) = C(0)\exp\{-t((Q/V) + K_C)\} \quad (10.64)$$

In this case one can solve for the time t_α required for the constituent to reach a fraction

$(1 - \alpha)$ of the initial concentration $C(0)$ (i.e., $C(t)/C_e = 1 - \alpha$). The result is Eq. 10.62.

Equation 10.60 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:

$$\text{Minimize } X \quad (10.65)$$

The following mass balance and constituent concentration constraints apply for each period t :

$$\begin{aligned} C(t) = &[W_C(t)(1 - X)/(Q(t) + k_C V)] \\ &[1 - \exp\{-\Delta t((Q(t)/V) + k_C)\}] \\ &+ C(t-1)\exp\{-\Delta t((Q(t)/V) + k_C)\} \end{aligned}$$
$$(10.66)$$

$$C(t) \leq C(t)^{max} \quad (10.67)$$

If each period t is a within-year period, and if the waste loadings 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.

10.4.11.3 Stratified Impoundments

Many deep reservoirs and lakes become stratified during particular times of the year. Vertical temperature gradients arise that imply vertical density gradients. The depth-dependent density gradients effectively prevent complete vertical mixing. Particularly in the summer season, two zones may form, 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 (Fig. 10.17).

Because of stratification many models divide the depth of water into layers, each of which is
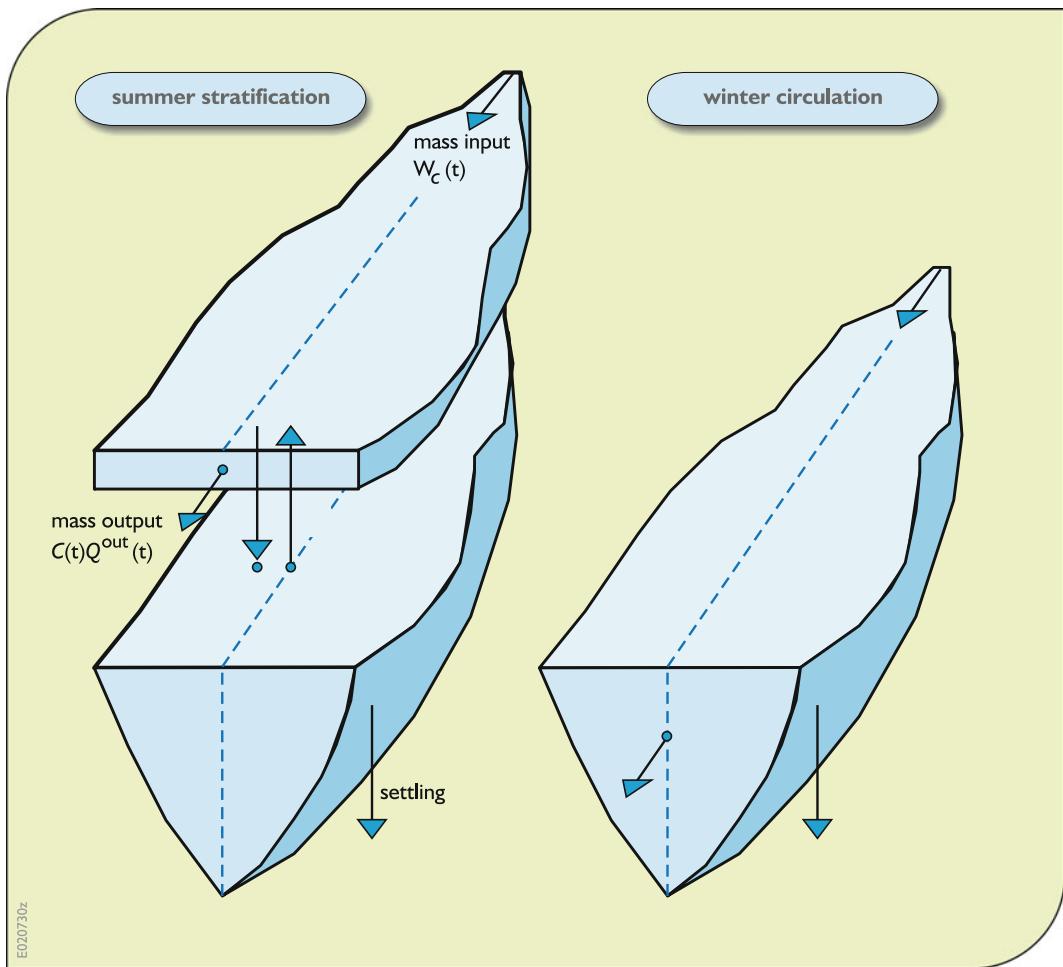
assumed to be fully mixed. To illustrate this approach without getting into too much detail, consider a simple two-layer lake in the summer that becomes a one-layer lake in the winter. This is illustrated in Fig. 10.18.

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 the 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, $kC_e(t)$:

$$dC_e(t)/dt = (1/V_e(t))\{(W_C(t) - C_e(t)Q^{out}(t)) + (v/D_T)[C_h(t)V_h(t) - C_e(t)V_e(t)]\} - kC_e(t)$$

$$(10.68)$$

$$dC_h(t)/dt = -kC_h(t) - (v/D_T)[C_h(t) - C_e(t)V_e(t)/V_h(t)] - sH_h(t)C_h(t)$$

$$(10.69)$$

Fig. 10.18 Lake stratification during summer and complete mixing during winter season

In the above two equations, V_e and V_h (m^3) are the time-dependent volumes of the epilimnion and hypolimnion, respectively, k (1/day) 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).

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

$$C(t) = C_e(t)V_e(t) + C_h(t)V_h(t)/[V_e + V_h] \tag{10.70}$$

$$dC(t)/dt = (1/V(t))\{(W_C(t) - C(t)Q^{out}(t)) - \text{decay } kC_e(t) - sH(t)C(t) \tag{10.71}$$

At the beginning of the summer season, each epilimnion and hypolimnion concentration will be the same.

$$C_e(t) = C(t) \tag{10.72}$$

$$C_h(t) = C(t) \tag{10.73}$$

10.5   Simulation Methods

Most who will be using water quality models will be using simulation models that are commonly available from governmental agencies (e.g., US EPA), universities, or private consulting and research institutions such as the Danish Hydraulics Institute, Wallilngford software or Deltares (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 incorporating 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 hydrologic or hydrodynamic model calculates the flow of water between all the segments and volume elements. In each simulation time step plug flow enters these segments or volume elements from upstream segments or elements. Flow also exits these segments or volume elements to downstream segments or elements. During this time the constituents can decay or grow, as appropriate, depending 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.

10.5.1   Numerical Accuracy

As presented in Sect. 10.4, equations describing water quality processes typically include time rate of change terms such as dC/dt. While it is possible to solve analytically some of these differential equations, 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 modeler because of these numerical methods.

Consider first the relationship between the stream, river, or lake segments and the duration of time steps, Δt. 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 (1/day), is

$$dC/dt = -kC \tag{10.74}$$

The finite difference approximation of this equation can be written

$$C(t + \Delta t) - C(t) = -C(t)k\Delta t \tag{10.75}$$

or

$$C(t + \Delta t) = C(t)(1 - k\Delta t) \qquad (10.76)$$

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 10.19 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 Eq. 10.110, namely

$$C(t + \Delta t) = C(t)\exp\{-k\Delta t\} \qquad (10.77)$$

Reducing the value of Δt will increase the accuracy of the numerical solution. Any value of Δt can be divided by a positive integer n to become $1/n$th of its original value. In this case the predicted concentration $C(t + \Delta t)$ will equal

$$C(t + \Delta t) = C(t)(1 - k\Delta t/n)^n \qquad (10.78)$$

For example if $k\Delta t = 1$, and $n = 2$, the final concentration ratio will equal

$$C(t + \Delta t)/C(t) = (1 - 1/2)^2 = 0.25 \quad (10.79)$$

Compare this 0.25–0.37, 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 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 less than the actual value based on Eq. 10.77, 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 Δt exceeds n/k. In any event, the predicted concentrations are not very useful.

Letting $m = -n/k\Delta t$, Eq. 10.78 can be written as
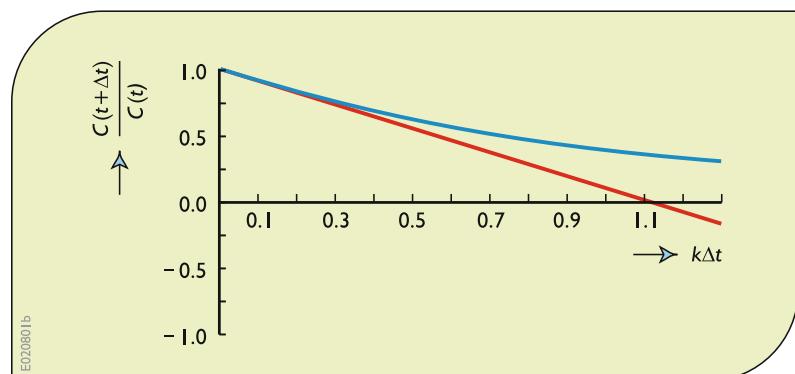
$$C(t + \Delta t) = C(t)(1 + 1/m)^{m(-k\Delta t)} \qquad (10.80)$$

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, Eq. 10.80 becomes Eq. 10.77, the exact solution to Eq. 10.74.

10.5.2 Traditional Approach

Most water quality simulation models simulate quality over a period of time. Time is divided into discrete intervals and the water and wastewater flows are assumed constant within each of those time period intervals. Each water body is divided into segments or volume elements and these "computational cells" 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

Fig. 10.19 Plot of numerical approximation (*red line*) based on Eq. 10.77 compared to the true analytical (*blue line*) value obtained from Eq. 10.76
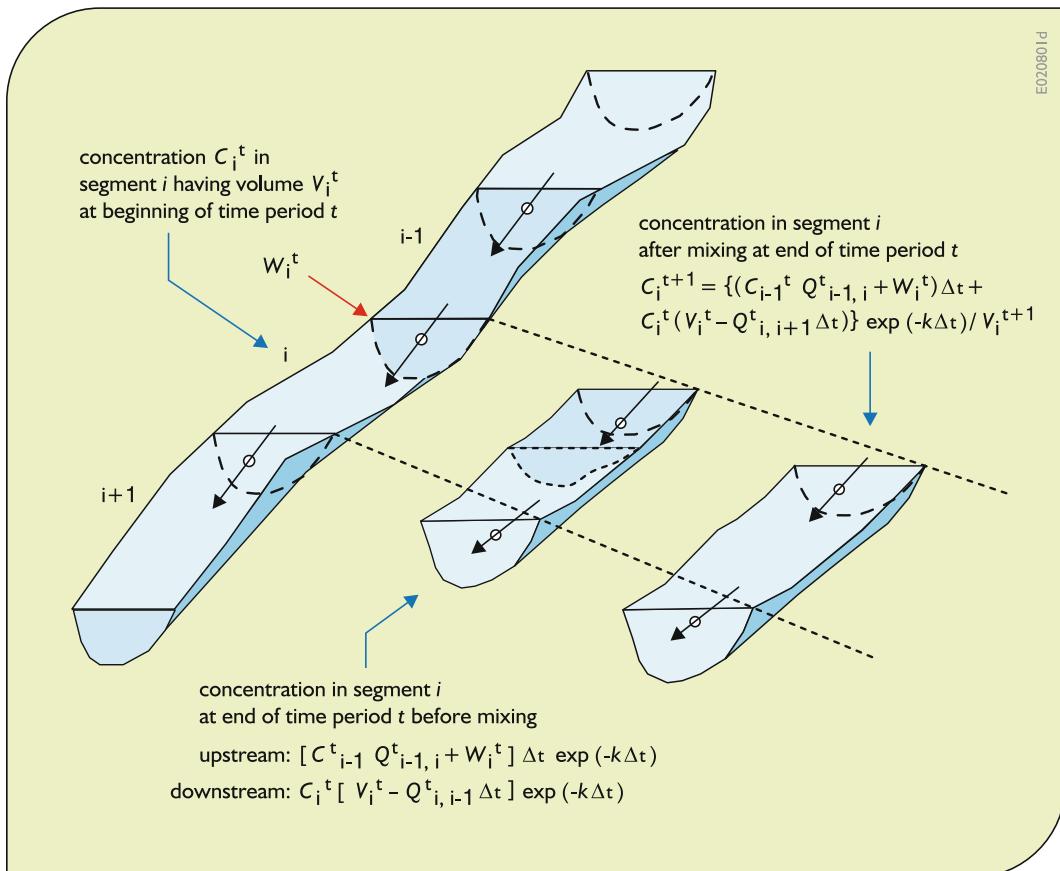
of the next time step. The larger the computational cell the greater is the dispersion.

This method is illustrated in Fig. 10.20. 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 Δt 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 Δt 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 its 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 Fig. 10.20, 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 shown on the left represents a portion of a stream or river divided into well-mixed segments. During a period t waste constituents enter reach segment

Fig. 10.20 Water quality modeling approach showing a water system schematized into reach segments or 'computational cells'

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 wastes, 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, Δx_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 increasing flow velocities, also reduces the allowable duration of each time period t. The duration of each simulation time step Δt 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

$$\Delta t \leq T_i \qquad (10.81)$$

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 Δx_i, cross-section area A_i and average flow Q_{it}, the restriction is

$$\Delta t \leq \min\{\Delta x_i A_i / Q_{it}; \forall i, t\} \qquad (10.82)$$

If time steps are chosen which violate this condition, then numerical solutions will be in error. The restriction defined by Eq. 10.82 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 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.
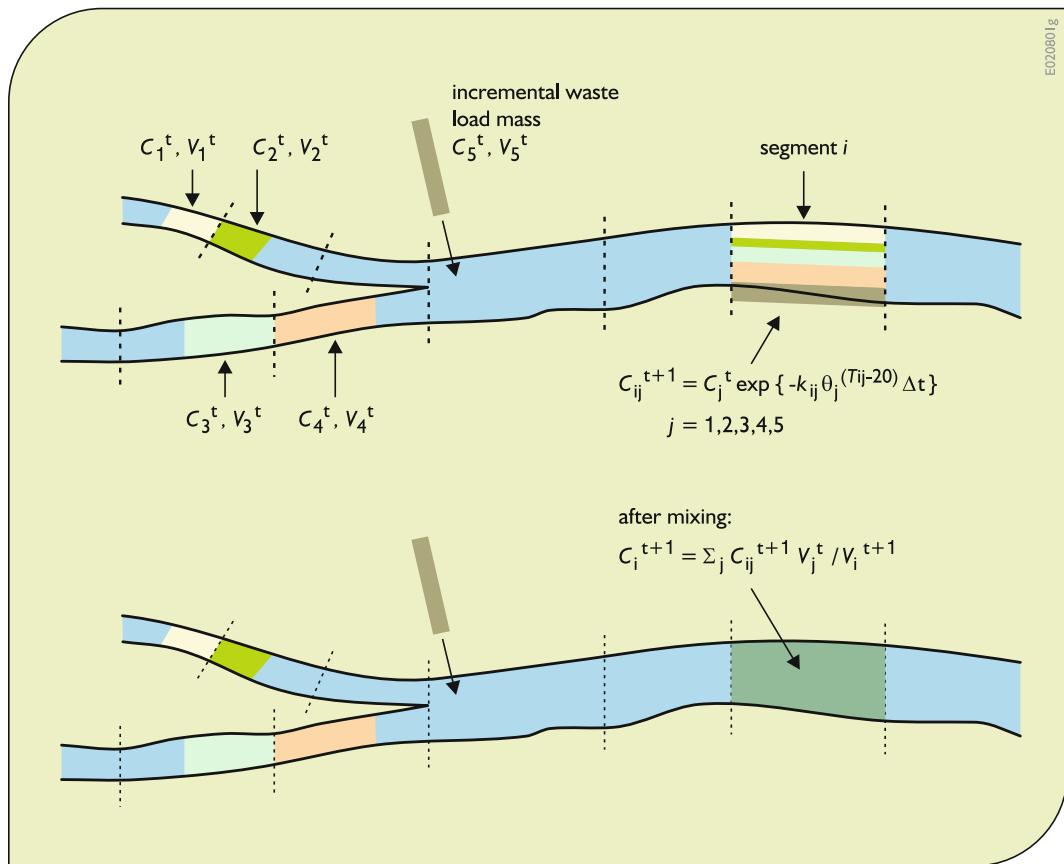
10.5.3 Backtracking Approach

An alternative backtracking approach to water quality simulation eliminates the need to consider the simulation time step duration restriction indicated by Eq. 10.82 (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 having a time interval Δt. The contaminants could come from segments in the same river reach or storage site, or from upstream river reach 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 Δt, as shown in Fig. 10.21.

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 Fig. 10.21, 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

Fig. 10.21 The backtracking approach for computing the concentrations of constituents in each reach segment or volume element i during time step duration of Δt

positions of all reach intervals that will be at the corresponding boundaries of segment i at the end of time period t. This requires computing the velocities through each of the intermediate segments or volume elements. Second, the changes in the amounts of the modeled quality constituents, e.g., temperature, organics, nutrients and toxics, are calculated assuming plug flow during the time interval, Δt, 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 or greater than Δt. 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 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 10.21 illustrates an example of backtracking involving simple first-order decay processes. Assume contaminants that end up in reach segment i at time 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 Δt in each segment or volume element is determined by the following differential equation:

$$dC_j^t/dt = -k_j\theta_j^{(T-20)}C_j^t \qquad (10.83)$$

The decay rate constant k_j, temperature correction coefficient θ_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 water travels during the period t from sites j to segment i.

Integrating Eq. 10.83 yields:

$$C_j^{t+1} = C_j^t\exp\{-k_j\theta_j^{(T-20)}\Delta t\} \qquad (10.84)$$

Since Δt 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}.

$$C_{ij}^{t+1} = C_j^t\exp\{-k_{ij}\theta_j^{(Tij-20)}\Delta t\} \qquad (10.85)$$

In Eq. 10.85 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 $t + 1$.

10.5.4 Model Uncertainty

There are two significant sources of uncertainty in water quality management models. One stems from incomplete knowledge or lack of sufficient data needed 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 socioeconomic 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 modeling capabilities.

However, even if it were possible to eliminate knowledge uncertainty, complete model prediction certainty in support of water quality management decisions will likely never be achieved until we can predict the variability of natural processes. This is the other significant source of uncertainty in water quality management models. 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, i.e., 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 both types of uncertainty, knowledge uncertainty, and unpredictable variability or randomness, cannot 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 actions taken to meet the desired water quality criteria and designated uses of water bodies. Chapters 6, 7, and 8 suggests some ways of characterizing this uncertainty.

10.6 Conclusions—Implementing a Water Quality Management Policy

This chapter provides only a brief introduction to some of the relationships contained in water quality models. As can be said for other chapters as well, entire texts, and very good ones, have been written on this subject (see, for example, Chapra and Reckhow 1983; Chapra 1997; McCutcheon 1989; Orlob 1983; Schnoor 1996; Thomann and Mueller 1987). Water quality modeling and management require skill and data. Skill comes with experience.

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. Adaptive approaches strive toward achieving water quality standards while relying on monitoring and experimentation to reduce uncertainty. This is often a way one can proceed given the complexity of the real world compared to the predictive models and data and time usually available at the time a water quality analysis is needed. Starting with simple analyses and iteratively expanding data collection and modeling 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 regulatory rules 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 modeling.

Regardless of what immediate actions are taken, there may not be an immediate measurable response. For example there may be significant time lags between 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 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.

Waste load allocations will inevitably be required if quality standards are not being met. These allocations involve costs. Different allocations will have a 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 US 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 fulfillment of other objectives must have made such a policy politically acceptable, even though expensive. However, more than these types of waste load allocations policies will be needed for many of

the ecosystem restoration efforts that are increasingly being made. Restoration activities are motivated in part by a recognition of the services ecosystems provide for water quality management.

Our capabilities of including 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, 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 assure that an action leading to attainment of designated uses will be initially identified. Adaptive management is a reasonable option in most cases for allowing water quality management programs to move forward in the face of considerable uncertainties.

References

Ambrose, R. B., Barnwell, T. O., McCutcheon, S. C., & Williams, J. R., (1996). Computer models for water quality analysis. In L. W. Mays (Ed.), *Water resources handbook*. New York: McGraw-Hill.

Borsuk, M. E., Stow, C. A., Higdon, D., & Reckhow, K. H. (2001). *A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones. Ecological modeling*. Durham, NC: Duke University.

Brown, L. C., & 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., & Cole, T. (1995). *User's guide to the CE-QUAL-ICM three-dimensional eutrophication model, release version 1.0*. Technical Report EL-95-15, US Army Eng. Waterways Experiment Station, Vicksburg, MS.

Chapra, S. C. (1997). *Surface water-quality modeling* (p. 844). New York: McGraw-Hill.

Chapra, S. C., & Reckhow, K. H. (1983). *Engineering approaches for lake management* (Vol. 2, Mechanistic modeling). Ann Arbor, MI: Ann Arbor Press, Inc.

DeMarchi, C., Ivanov, P., Jolma, A., Masliev, I., Smith, M., & Somlyódy, L. (1999). Innovative tools for water quality management and policy analysis: DESERT and STREAMPLAN. *Water Science and Technology, 40*(10), 103–110.

Elmore, H. L., & Hayes, T. W. (1960). Solubility of atmospheric oxygen in water. *Jouranl of Environmental Engineering Division ACSE, 86*(SA4), 41–53.

Ivanov, P., Masliev, I., De Marchi, C., & Somlyódy, L. (1996). *DESERT—Decision support system for evaluating river basin strategies, user's manual*. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Karr, J. R. (1990). *Bioassessment and non-point source pollution: An overview* (pp 4-1 to 4-18 in second national symposium on water quality assessment). Washington, DC: EPA Office of Water.

Manson, J. R., & Wallis, S. G. (2000). A conservative semi-lagrangian fate and transport model for fluvial systems—I. Theoretical development. *Water Resources, 34*(15), 3769–3777.

McCutcheon, S. C. (1989). *Water quality modeling* (Vol. 1). Boca Raton, FL: CRC Press.

National Research Council (NRC). (1992). *Restoration of aquatic ecosystems*. Washington, D.C.: National Academy Press.

National Research Council (NRC). (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* (122 p). Washington, D.C.: Water Science and Technology Board, National Academy Press.

Orlob, G. T. (Ed.). (1983). *Mathematical modeling of water quality: Streams, lakes and reservoirs*. Chichester, UK: Wiley.

Reichert, P. (1994). AQUASIM—A tool for simulation and data analysis of aquatic systems. *Water Science and Techonolgy, 30*(2), 21–30.

Schnoor, J. L. (1996). *Environmental modeling, fate and transport of pollutants in water, air and soil*. New York: Wiley.

Stewart, T. R. (2000). Uncertainty, judgment, and error in prediction. In D. Sarewitz, R. A. Pielke Jr., & R. Byerly Jr. (Eds.), *Prediction: Science, decision making, and the future of nature*. Washington, D.C.: Island Press.

Streeter, H. W., & 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.: U.S. Public Health Service.

Thomann, R. V., & Mueller, J. A. (1987). *Principles of surface water quality modeling and control*. New York: Harper & Row Publishers.

USEPA. (2001). *BASINS version 3.0 user's manual. EPA-823-B-01-001* (337 p.). Washington, DC: EPA Office of Water and Office of Science and Technology.

WL|Delft Hydraulics. (2003). *Delft3D-WAQ, user manual; Versatile water quality modeling in 1D, 2D, or 3D systems including physical, (bio)chemical and biological processes.*

Yin, H. (2002). *Development of a watershed information system, ms thesis.* Ithaca, NY: Civil and Environmental Engineering, Cornell University.

Additional References (Further Reading)

ASCE. (1999). National stormwater best management practices (bmp) data-base. Version 1.0. 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, Washington, DC. User's Guide and CD.

Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water Resources Research, 23*, 1393–1442.

Beck, M. B., & van Straten, G. (Eds.). (1983). *Uncertainty and forecasting of water quality.* Berlin: Springer.

Biswas, A. K. (Ed.). (1997). *Water resources: Environmental planning, management, and development.* New York: The McGraw-Hill Companies Inc.

Bowie, G. L., Mills, W. B., Porcella, D. B., Campbell, C. L., Pagenkopf, J. R., Rupp, G. L., et al. (1985). *Rates, constants, and kinetics formulations in surface water quality modeling* (2nd edn). Athens, GA: Report EPA/600/3-85/040, US EPA.

Churchill, M. A., Elmore, H. L., & Buckingham, R. A. (1962). Prediction of stream reaeration rates. *Journal of the Sanitary Engineering Division ASCE* SA4.1, Proc. Paper 3199.

Delvigne, G. A. L. (1980). *Natural reaeration of surface water.* WL|Delft Hydraulics, Report on literature study R1149 (in Dutch), Delft, NL.

DiToro, D. M., Paquin, P. R., Subburamu, K., & Gruber, D. A. (1990). Sediment oxygen demand model: Methane and ammonia oxidation. *Journal of Environmental Engineering, ACSE 116*(5), 945–986.

Dobbins, W. E. (1964). BOD and oxygen relationships in streams. *Journal of the Sanitary Engineering Division, 90*(SA3), 53–78.

EU. (2001a). The EU water framework directive. http://europa.eu.int/water/water-framework/index_en.html. Accessed 21 June 2001.

EU. (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 2001.

EU. (2001c). Water protection and management, urban waste water treatment. http://www.europa.eu.-int/scadplus/leg/en/lvb/l28008.htm. Accessed 20 June 2001.

Engelund, F., & Hansen, E. (1967). *A monograph on sediment transport in alluvial streams.* Copenhagen, DK: Teknisk forlag.

Fagerbakke, K. M., Heldal, M., & Norland, S. (1996). Content of carbon, nitrogen, oxygen, sulfur and phosphorus in native aquatic and cultured bacteria. *Aquatic Microbial Ecology, 10*, 15–27.

Gromiec, M. J., Loucks, D. P., & Orlob, G. T. (1982). Stream quality modeling. In G. T. Orlob (Ed.), *Mathematical modeling of water quality.* Chichester, UK: Wiley.

Hornberger, G. M., & Spear, R. C. (1981). An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management, 12*, 7–18.

Hornberger, G. M., & Spear, R. C. (1983). An approach to the analysis of behavior and sensitivity in environmental systems. In M. B. Beck & G. van Straten (Eds.), *Uncertainty and forecasting of water quality* (pp. 101–116). Berlin: Springer.

Karr, J. R. (2000). Health, integrity, and biological assessment: The importance of whole things. In D. Pimentel, L. Westra, & R. F. Noss, (Eds.), *Ecological integrity: integrating environment, conservation, and health* (pp. 209–226, 214–215). Washington, DC: Island Press.

Karr, J. R., & Dudley, D. R. (1981). Ecological perspective on water quality goals. *Environmental Management, 5*, 55–68.

Karr, J. R., & Chu, E. W. (2000). Sustaining living rivers. *Hydrobiologia, 422*(423), 1–14.

Krone, R. B. (1962). *Flume studies of the transport of sediment in estuarial shoaling processes.* Berkeley: University of California, Hydraulic and Sanitary Engineering Laboratory.

Langbien, W. B., & Durum, W. H. (1967). *The aeration capacity of streams* (p. 542). Washington, D.C.: USGS.

Los, F. J. (1991). *Mathematical Simulation of algae blooms by the model BLOOM II, Version 2, T68.* WL|Delft Hydraulics Report.

Los, F. J. et al. (1992). *Process formulations DBS.* WL|Delft Hydraulics, Model documentation T542 (in Dutch), Delft, NL.

Maidment, D. R. (Ed.). (1993). *Handbook of hydrology.* New York: McGraw-Hill.

Masliev, I., Somlyódy, L., & Koncsos, L. (1995). *On reconciliation of traditional water quality models and activated sludge models.* Working Paper WP 95–18, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Mills, W. B., Porcella, D. B., Ungs, M. J., Gherini, S. A., Summers, K. V., Mok, L., et al. (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.

Monbaliu, J., Hargreaves, J. C., Carretero, J.-C., Gerritsen, H., & Flather, R. A. (1999). Wave modelling in the PROMISE project. *Coastal Engineering, 37*(3–4), 379–407.

Morgan, M. G., & Henrion, M. (1990). *Uncertainty* (p. 332). New York: Cambridge University Press.

National Academy of Public Administration. (2000). *Transforming environmental protection for the 21st 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 AWRA, 35*(4), 717–727.

Novotny, V., & Olem, H. (1994). Water quality: Prevention, identification and management of diffuse pollution. New York: Van Nostrand-Reinhold (distributed by Wiley, Ohio EPA).

O'Connor, D. J. (1961). Oxygen balance of an estuary. *Journal of Sanitary Engineering Division ASCE, 86* (SA3), 35–55.

O'Connor, D. J., & Dobbins, W. E. (1958). Mechanism of reaeration in natural streams. *Transactio of ASCE, 123*, 641–666.

Owens, M., Edwards, R. W., & Gibbs, J. W. (1964). Some reaeration studies in streams. *International Journal of Air and Water Pollution, 8*, 469–486.

Partheniades, E. (1962). *A study of erosion and deposition of cohesive soils in salt water*. Berkeley: University of California.

Peters, R. H. (1991). *A critique for ecology* (p. 366). 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, 10*(3), 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, 43*(5), 51–60.

Reichert, P., Borchardt, D., Henze, M., Rauch, W., Shanahan, P., Somlyódy, L., et al. (2001a). River water quality model no. 1 (RWQM1): II. Biochemical process equations. *Water Science and Technology, 43* (5), 11–30.

Reichert, P., & Vanrolleghem, P. (2001). Identifiability and uncertainty analysis of the River Water Quality Model No. 1 (RWQM1). *Water Science and Technology 43*(7), 329–338.

Roesner, L. A., Giguere, P. R., & Evenson, D. E. (1981). *Computer program documentation for the stream quality model QUAL–II*. Report EPA 600/9–81–014, US EPA, Athens, GA.

Shanahan, P., Borchardt, D., Henze, M., Rauch, W., Reichert, P., Somlyódy, L., et al. (2001). River water quality model no. 1 (RWQM1): I Modeling approach. *Water Science and Technology, 43*(5), 1–9.

Shanahan, P., Henze, M., Koncsos, L., Rauch, W., Reichert, P., Somlyódy, L., et al. (1998). River water quality modeling: II. Problems of the art. *Water Science and Technology, 38*(11), 245–252.

Shen, H. W., & Julien, P. Y. (1993). *Erosion and sediment transport, chapter 12 of handbook of hydrology* (D. R. Maidment Ed.). New York: McGraw-Hill, Inc.

Smith, R. A., Schwarz, G. E., & Alexander, R. B. (1997). Regional interpretation of water-quality monitoring data. *Water Resources Research, 33*(12), 2781–2798.

Smits, J. (2001). *DELWAQ-BLOOM-switch*. Delft, NL: Delft Hydraulics.

Somlyódy, L. (1982). Water quality modeling: A comparison of transport-oriented and ecology-oriented approaches. *Ecological Modelling, 17*, 183–207.

Somlyódy, L., & van Straten, G. (Eds.). (1986). *Modeling and managing shallow lake eutrophication, with application to lake balaton*. Berlin: Springer.

Somlyódy, L., & Varis, O. (1993). Modeling the quality of rivers and lakes. In P. Zannetti (Ed.), *Environmental modeling: Computer methods and software for environmental modeling: Computer methods and software for simulating environmental pollution and its adverse effects* (Vol. 1, pp. 213–257). Southampton, UK: CMP.

Somlyódy, L., Henze, M., Koncsos, L., Rauch, W., Reichert, P., Shanahan, P., et al. (1998). River water quality modeling: III. Future of the art. *Water Science and Technology, 38*(11), 253–260.

Spanou, M., & Chen, D. (2001). Water-quality modeling of the upper Mersey River system using an object-oriented framework. *Journal of Hydroinformatics, 3*(3), 173–194.

Spear, R., & Hornberger, G. M. (1980). Eutrophication in peel inlet—II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research, 14*, 43–49.

Stumm, W., & Morgan, J. J. (1981). *Aquatic chemistry*. New York: Wiley.

Thomas, H. A., Jr. (1948). Pollution load capacity of streams. *Water Sewage Works, 95*, 409.

Toxopeus A. G. (1996). *An interactive spatial and temporal modeling system as a tool in ecosystem management*. International Institute for Aerospace Survey and Earth Sciences. ITC.

Ulanowicz, R. E. (1997). *Ecology, the ascendant perspective* (p. 201). 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: Second edition*. 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.: US EPA.

USEPA. (1998). *Lake and reservoir bioassessment and biocriteria: Technical guidance document*. EPA 841-B-98-007. Washington, D.C.: EPA Office of Water.

USEPA. (1999). *Draft guidance for water quality-based decisions: The TMDL process* (2nd ed.). 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 der Molen, D. T., Los, F. J., van Ballegooijen, L., van der Vat, M. P. (1994). Mathematical modelling as a tool for management in eutrophication control of shallow lakes. *Hydrobiologia, 275/276*, 479–492.

Van Pagee, J. A. (1978). Natural reaeration of surface water by the wind. WL|Delft Hydraulics, Report on literature study R1318-II (in Dutch).

Van Rijn, L. C. (1984). Bed load transport (part I), suspended load transport (part II). *Journal of Hydraulic Engineering, 110*(10, 11), 1431–1456, 1613–1641.

Van Straten, G. (1983). Maximum likelihood estimation of parameters and uncertainty in phytoplankton models. In M. B. Beck & G. van Straten (Eds.), *Uncertainty and forecasting of water quality*. Berlin: Springer.

Young, P. (1998). Data-based mechanistic modeling of environmental, ecological, economic and engineering systems. *Environmental Modeling & Software, 13*, 105–122.

Young, W. J., Lam, D. C. L., Ressel, V., & Wong, J. W. (2000). Development of an environmental flows decision support system. *Environmental Modeling & Software, 15*, 257–265.

Exercises

1. The common version of the Streeter–Phelps equations for predicting biochemical oxygen demand BOD and DO deficit D concentrations are based on the following two differential equations

  (a) $\mathrm{d(BOD)}/\mathrm{d}t = -K_d(\mathrm{BOD})$

  (b) $\mathrm{d}D/\mathrm{d}t = K_d\,(\mathrm{BOD}) - K_aD,$

where K_d is the deoxygenating rate constant (T^{-1}), K_a is the reaeration-rate constant (T^{-1}), and τ is the time of flow along a uniform reach of stream in which dispersion is not significant. Show the integrated forms of (a) and (b).

10.2 Based on the integrated differential equations in Exercise 10.1

  (a) Derive the equation for the distance X_c downstream from a single point source of BOD that for a given streamflow will have the lowest dissolved oxygen concentration.

  (b) Determine the relative sensitivity of the deoxygenation rate constant K_d and the reaeration rate constant K_a on the critical distance X_c and on the corresponding critical deficit D_c. For initial conditions, assume that the reach has a velocity of 2 m/s (172.8 km/day), a K_d of 0.30 per day, and a K_a of 0.4 per day. Assume that the DO saturation concentration is 8 mg/l, the initial deficit is 1.0 mg/l, and the BOD concentration at the beginning of the reach (including that discharged into the reach at that point) is 15 mg/l.

10.3 To account for settling of BOD, in proportion to the BOD concentration, and for a constant rate of BOD addition R due to runoff and scour, and oxygen production ($A > 0$) or reduction ($A < 0$) due to plants and benthal deposits, the following differential equations have been proposed:

  (a) $\mathrm{d(BOD)}/\mathrm{d}\tau = -(K_d + K_s)\,(\mathrm{BOD}) + \mathrm{R}$
  $$\tag{10.86}$$

  (b) $\mathrm{d}D/\mathrm{d}\tau = \mathrm{K_d(BOD)} - \mathrm{K_aD} - \mathrm{A},$
  $$\tag{10.87}$$

where K_s is the settling rate constant (T^{-1}) and τ is the time of flow. Integrating these two equations results in the following deficit equation:

$$D_\tau = \frac{K_d}{K_a - (K_d + K_s)}\left[\left(BOD_0 - \frac{R}{K_d + K_S}\right)\right.$$
$$\{\exp[K_d + K_S] - \exp(-K_a\tau)\}]$$
$$+ \frac{K_d}{K_a}\left\{\left(\frac{R}{K_d + K_s} - \frac{A}{K_d}\right)[1 - \exp(-K_a\tau)]\right\}$$
$$\left. + D_0\exp(-K_a\tau)\right.$$
$$\tag{10.88}$$

where BOD$_o$ and D_o are the BOD and DO deficit concentrations at $\tau = 0$.

(a) Compare this equation with that found in Exercise 10.1 if Ks, R, and A are 0
(b) Integrate Eq. (10.86) to predict the BOD$_\tau$ at any flow time τ.

10.4 Develop finite difference equations for predicting the steady-state nitrogen component and DO deficit concentrations D in a multi-section one-dimensional estuary. Define every parameter or variable used.

10.5 Using Michaelis–Menten kinetics develop equations for

(a) Predicting the time rate of change of a nutrient concentration N (dN/dt) as a function of the concentration of bacterial biomass B;
(b) Predicting the time rate of change in the bacterial biomass $B(dB/dt)$ as a function of its maximum growth rate μ_B^{max}, temperature T, B, N, and the specific-loss rate of bacteria ρ_B; and
(c) Predicting the time rate of change in DO deficit (dD/dt) also as a function of N, B, ρ_B, and the reaeration-rate constant $K_a(T^{-1})$.

How would these three equations be altered by the inclusion of protozoa P that feed on bacteria, and in turn require oxygen? Also write the differential equations for the time rate of change in the concentration of protozoa $P(dP/dt)$.

10.6 Many equations for predicting stream temperature use Eulerian coordinates. The actual behavior of the stream temperature is more easily demonstrated if Lagrangian coordinates (i.e., time of flow t rather than distance X) are used. Assuming insignificant dispersion, the "time-of-flow" rate of temperature change of a water parcel as it moves downstream is

$$dT/d\tau = \lambda(T_E - T)/\rho cD$$
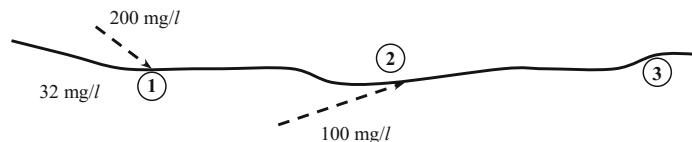
(a) Assuming that λ, D, and T_E are constant over interval of time of flow $t_2 - t_1$, integrate the equation above to derive the temperature T_1 at locations X_1.
(b) Develop a model for predicting the temperature at a point in a nondispersive stream downstream from multiple point sources (discharges) of heat.

10.7 Consider three well-mixed bodies of water that have the following constant volumes and freshwater inflows

Water body	Volume (m^3)	Flow (m^3/s)	Displacement time
1	3×10^{12}	3×10^3	3.17 years
2	3×10^8	3×10^2	11.6 days
3	3×10^4	3×10^4	2.8 h

The first body is representative of the Great Lakes in North America, the second is characteristic in size to the upper New York harbor with the summer flow of the Hudson River, and the third is typical of a small bay or cove. Compute the time required to achieve 99% of the equilibrium concentration, and that concentration, of a substance having an initial concentration, and that concentration, of a substance having an initial concentration of 0 (at time = 0) and an input of N (M T^{-1}) for each of the three water bodies. Assume that the decay rate constant K is 0, 0.01, 0.05, 0.25, 1.0, and 5.0 days^{-1} and compare the results.

10.8 Consider the water pollution problem as shown in Figure below. There are two sources of nitrogen, 200 mg/l at site 1 and 100 mg/l at site 2, going into the river, whereas the nitrogen concentration in the river just upstream of site 1 is 32 mg/l.

The unknown variables are the fraction of nitrogen removal at each of those sites that would achieve concentrations no greater than 20 and 25 mg/l just upstream of site 2 and at site 3, respectively, at a total minimum cost. Let those nitrogen removal fractions be X_1 and X_2. Assuming unit costs of removal as \$30 and \$20 at site 1 and site 2, respectively, the model can be written as

Minimize $30X_1 + 20X_2$
Subject to : $200(1 - X_1)0.25 + 8 \leq 20$
              $200(1 - X_1)0.15 + 100(1-X_2)0.60 + 5 \leq 25$
              $X_1 \leq 0.9, X_2 \leq 0.9$
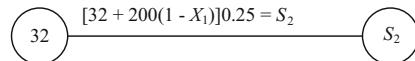
Another way to write the two quality constraints of this model is to define variables S_i ($i = 1, 2, 3$) as the concentration of nitrogen just upstream of site i. Beginning with a concentration of 32 mg/l just upstream of site 1, the concentration of nitrogen just upstream of site 2 will be

$$[32 + 200(1 - X_1)]0.25 = S_2 \text{ and } S_2 \leq 20.$$

The concentration of nitrogen at site 3 will be

$$[S_2 + 100(1 - X_2)]0.60 = S_3 \text{ and } S_3 \leq 25.$$

This makes the problem easier to solve using discrete dynamic programming. The nodes or states of the network can be discrete values of S_i, the concentration of nitrogen in the river at sites i (just upstream of sites 1 and 2 and at site 3). The links represent the decision variable values, X_i that will result in the next discrete concentration, S_{i+1} given S_i. The stages i are the different source sites or river reaches. A section of the network in stage 1 (reach from site 1 to site 2) will look like

So if S_2 is 20, X_1 will be 0.76; if S_2 is 15, X_1 will be 0.86. For S_2 values of 10 or less X_1 must exceed 0.90 and these values are infeasible. The cost associated with the link or decision will be 30 X_1.

Setup the dynamic programming network. It begins with a single node representing the state (concentration) of 32 mg/l just upstream of site 1. It will end with a single node representing the state (concentration) 25 mg/l. The maximum possible state (concentration value just upstream of site 2 must be no greater than 20 mg/l. You can use discrete concentration values in increments of 5 mg/l. This will be a very simple network. Find the least-cost solution using both forward and backward moving dynamic programming procedures. Please show your work.

10.9 Identify three alternative sets (feasible solutions) of storage lagoon volume capacities V and corresponding land application areas A and irrigation volumes Q_{2t} in each month t within a year that satisfy a 10 mg/l maximum NO$_3$–N content in the drainage water of a land disposal system. In addition to the data listed below, assume that the influent nitrogen n_{1t} is 50 mg/l each month, with 10% ($\alpha = 0.1$) of the nitrogen in organic form. Also assume that the soil is a well-drained silt loam containing 4500 kg/ha of organic nitrogen in the soil above the drains. The soil has a monthly drainage capacity d of 60 cm and has a

field capacity moisture content M of 10 cm. Maximum plant nitrogen uptake values N_t^{max} are 35 kg/ha during April till October, and 70 kg/ha during May till September. Finally, assume that because of cold temperatures, no wastewater irrigation is permitted during November till March. December, January, and February's precipitation is in the form of snow and will melt and be added to the soil moisture inventory in March.

removal is required at each discharge site, solve for the least-cost solution given the data in the accompanying table. Can you identify more than one type of model to solve this problem? How would this model be expanded to specifically include both carbonaceous BOD and nitrogenous BOD and non-point waste discharges?

Reach no.	Design BOD load (mg/l)	Present % removal load	Annual costs of various design BOD removal			
			60%	75%	85%	90%
1	248	67	0	22,100	77,500	1,20,600
2	408	30	6,30,000	7,80,000	9,87,000	11,70,000
3	240	30	2,10,000	2,77,500	3,23,000	3,78,000
4	1440	30	4,13,000	5,23,000	6,26,000	6,98,000
6	2180	30	5,00,000	6,38,000	7,90,000	9,00,000
7	279	30	8,40,000	10,72,000	12,32,500	13,50,000

Reach no.	Time of flow (days)	Wastewater discharge (10^3 m³/day)	Entering reach flow (10^3 m³/day)	Total reach flow (10^3 m³/day)	DO saturation conc. (mg/l)	Maximum allowable DO deficit (mg/l)	DO deficit of wastewater (mg/l)	DO Conc at beginning of reach (mg/l)	BOD Conc at beginning of reach (mg/l)	Av. deoxgn rate constant for reach (days⁻¹)	Reaeration rate constant days⁻¹
1	0.235	19	5129	5148	10.2	3.2	1	9.5	1.66	0.31	1.02
2	1.330	140	4883	5023	9.95	2.45	1	8	0.68	0.41	0.6
3	1.087	30	10,171	10,201	9	2	1	?	?	0.36	0.63
4	2.067	53	1120	1173	9.7	3.75	1	9.54	1	0.35	0.09
5	0.306	0	11,374	11,374	9	2.5	–	?	?	0.34	0.72
6	1.050	98	11,374	11,472	8.35	2.35	1	–	–	0.35	0.14
7	6.130	155	11,472	11,627	8.17	4.17	1	–	–	0.3	0.02

10.10 Consider the problem of estimating the minimum total cost of waste treatment in order to satisfy quality standards within a stream. Let the stream contain seven homogenous reaches r, reach $r = 1$ being at the upstream end and reach $r = 7$ at the downstream end. Reaches $r = 2$ and 4 are tributaries entering the mainstream at the beginning of 1, 3, 5, 6, and 7. Point sources of BOD enter the stream at the beginning of reaches 1, 2, 3, 4, 6, and 7. Assuming that at least 60% BOD

10.11 Discuss what would be required to analyze flow augmentation alternatives in Exercise 12.8. How would the costs of flow augmentation be defined and how would you modify water quality models to include flow augmentation alternatives?

10.12 Develop a dynamic programming model to estimate the least-cost number, capacity, and location of artificial aerators to ensure meeting minimum allowable DO standards where they would otherwise be

violated during an extreme low-flow design condition in a nonbranching section of a stream. Show how wastewater treatment alternatives, and their costs, could also be included in the dynamic programming model.

10.13 Using the data provided, find the steady-state concentrations C_t of a constituent in a well-mixed lake of constant volume $30 \times 10^6 \, \mathrm{m}^3$. The production N_{ti} of the constituent occurs at three sites i, and is constant in each of four seasons in the year. The required fractions of constituent removal P_i at each site i are to be set so that they are equal at all sites i and the maximum concentration in the lake in each period t must not exceed 20 mg/l.

Period, t	Days in period	Flow, Qt (10^3 m^3/day)	Constituent decay rate, constant, Kt (days^{-1})
1	100	90	0.02
2	80	150	0.03
3	90	200	0.05
4	95	120	0.04

Constituent discharge site, i	Constituent production (kg/day)
1	38,000
2	25,000
3	47,000

10.14 Suppose that the solution of a model such as that used in Exercise 10.13, or measured data, indicated that for a well-mixed portion of a saltwater lake, the concentrations of nitrogen ($i = 1$), phosphorus ($i = 2$), and silicon ($i = 3$) in a

particular period t were 1.1, 0.1, and 0.8 mg/l, respectively. Assume that all other nutrients required for algal growth are in abundance. The algal species of concern are three in number and are denoted by $j = 1, 2, 3$. The data required to estimate the probable maximum algal bloom biomass concentration are given in the accompanying table. Compute this bloom potential for all k_i and k equal to 0, 0.8, and 1.0.

Parameter (algae species index j)	Parameter value		
	1	2	3
a_{1j} = mg N/mg dry wt of algae j	0.04	0.01	0.20
a_{2j} = mg P/mg dry wt of algae j	0.06	0.02	0.10
a_{3j} = mg Si/mg dry wt of algae j	0.08	0.01	0.03
D_j = morality and grazing rate constant (days^{-1})	0.6	0.4	0.20
d_j = morality rate constant, (days^{-1})	0.3	0.1	0.10
v = extinction reduction rate constant for dead algae, (days^{-1})	0.07	0.07	0.07
h_j^{max} = max. extinction coef. (m^{-1})	0.07	0.07	0.10
$\eta_j^{min} Z$ = min. extinction coef. (m^{-1})	0.01	0.03	0.03
η_j = increase in extinction coef. per unit increase in mg/l (g/m^3) of dry wt of species j (m^2/g)	0.05	0.164	0.04
Nutrient index i	*1*	*2*	*3*
Nutrient	N	P	Si
μi = mineralization rate constant, (days^{-1})	0.02	0.69	0.62

River Basin Modeling

<div style="text-align: right">

11

</div>

Multipurpose river basin development typically involves the identification and use of both structural and nonstructural measures designed to increase the reliability and decrease the cost of municipal, industrial, and agriculture water supplies, to protect against droughts and floods, to improve quality, to provide for commercial navigation and recreation, to enhance aquatic ecosystems, 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 treatment treatment plants along with their distribution and collection systems. Nonstructural measures may include land-use controls and zoning, flood warning and evacuation measures, and economic incentives that affect human behavior with regard to water and watershed use. Planning the development and management of water resource systems involves identifying just what and when and where structural or nonstructural measures are needed, the extent to which they are needed, and their combined economic, environmental, ecological, and social impacts. This chapter introduces some modeling approaches for doing this. Having just reviewed some water quality modeling approaches in the previous chapter, this chapter focuses on quantity management.

11.1 Introduction

Various types of models can be used to assist those responsible for planning and managing various components of river systems. These components include streams, rivers, lakes, reservoirs, and wetlands, and diversions to demand sites that could be within or outside the basin boundaries. Each of these components can be impacted by water management policies and practices. The management of any single component can impact the performance of other components in the basin. Hence, for the overall management of the water in river basin systems, a systems view is usually taken. This systems view requires the modeling of multiple interacting and interdependent components. These multicomponent models are useful for analyzing alternative designs and management policies for improving the performance of integrated river basin systems.

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 impact the quality of that water as well. However, unless water allocations allocations and uses are based on requirements for water quality, such as for the dilution of pollution, water quality does not normally affect water quantity. For this reason among others, it is common to separate discussions of water quantity

management from water quality management (Chap. 10). However, when attempting to predict the impacts of any management policy on both water quantity and quality, both water quantity and quality models are needed.

This chapter begins with a discussion of selecting appropriate model time periods that will depend on the issues being addressed as well as on the variability of the water supplies and demands. Discussed next are methods for estimating streamflows at various sites of interest throughout a basin based on gage (measured) flows at other sites. Following these discussions several methods are reviewed and compared for estimating reservoir storage requirements for water supplies. Model components are defined for withdrawals and diversions, and for reservoir storage. Reservoir storage can serve the needs over time for water supply, flood control, recreation, and hydroelectric power generation. Next flood control structures, such as levees and channel flow capacity improvements at potential flood damage in a river basin are introduced. These components are then combined into a multiple purpose multi-objective planning model for a hypothetical 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.

11.2   Model Time Periods

When analyzing 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. In optimization models, weekly, monthly, or seasonal flows are commonly used as opposed to daily flows. The shortest time period duration usually considered in optimization models developed for identifying and evaluating alternative water management plans and operating policies is one that is no less than the time water takes to flow from the upper

end of the applicable river basin to the lower end of the basin. In this case stream and river flows can be defined by simple mass balance or continuity equations. For shorter duration time periods flow routing may be required.

The actual length or duration of each within-year period defined in a model may vary from period to period. Modeled time period time periods need not be equal. Generally what is important is to capture in the model the needed capacities of infrastructure that are determined in large measure by the variation in supplies and demands. These variations should be captured in the model by appropriately selecting the number and duration of time periods. If say over a three-month period there is little variation in both water supplies and demands or for the purposes water serves, such as flood control, hydropower, or recreation, there is no need to divide that three-month period into multiple time periods.

Another important factor to consider in making a decision regarding the number and duration of time periods to include in any model is the purpose for which the model is to be used. Some analyses are concerned only with identifying designs designs and operating policies of various engineering projects for managing water resources at some fixed time (say a typical year) in the future. Multiple years of hydrological records are used, usually in simulation models, 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 that may be needed to meet various goals. These static analyses are not concerned with investment project scheduling or sequencing.

Dynamic planning models are used to estimate the impacts of changing conditions over time. These changes could include hydrological inputs, economic, environmental and other objectives, water demands, and design and operating parameters. As a result, dynamic models generally span many more years than do

static models, but they may have fewer within-year periods.

11.3   Streamflow Estimation

Water resource managers need estimates of streamflows at each site, where management decisions are being considered. These streamflow estimates can be based on the results of rainfall-runoff models or on measured historical flows at gage sites. For modeling alternative management policies, these gage-based flows at the sites of interest should be those that would have occurred under natural conditions. These are called naturalized flows that have been derived from measured flows or rainfall–runoff models and then adjusted to take into account any upstream regulation and diversions. Many gage flow values reflect actions such as diversions and reservoir releases that occurred upstream that altered the downstream flows, unless such upstream water management and use policies are to continue, these measured gage flows should be converted to unregulated or natural flows prior to their use in management models.

Assuming that unregulated streamflow data are available at gage sites, these data can be used to estimate the unregulated flows at sites where they are needed. These sites would include any place where diversions might occur or where reservoirs for regulating flows might be built.
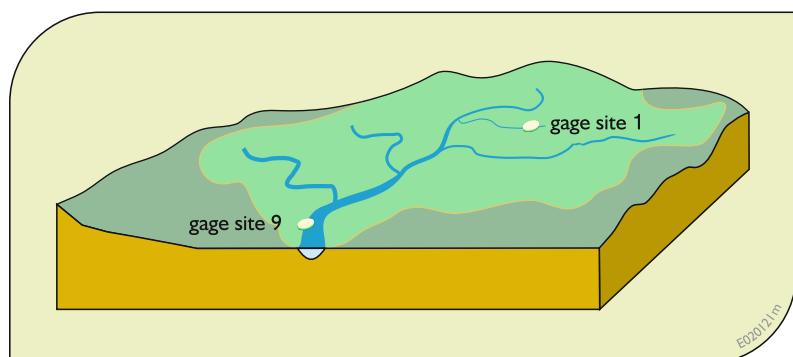
Consider, for example, the simple river basin illustrated in Fig. 11.1. Assume streamflows have been recorded over a number of years at gage sites 1 and 9. Knowledge of the flows Q_t^s in each period t at gage sites 1 and 9 permits the estimation of flows at any other site in the basin as well as the incremental flows between those sites in each period t.

The method used to estimate flows at ungaged sites will depend on the characteristics of the watershed or river basin. In humid regions where streamflows increase in the downstream direction due to rainfall runoff, 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 runoff per unit land area is typically assumed constant. In these situations, estimated flows, q_t^s, at any site s can be 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 gage sites.

For each gage site, the runoff per unit land area can be calculated by dividing the gage flow Q_t^s by the upstream drainage area, A^s. This can be done for each gage site in the basin. Thus for any gage 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 gage sites, such as illustrated in Fig. 11.1, the estimated streamflow at some ungaged site s can be a weighted combination of those unit area runoffs times the area contributing to the flow at site s. The nonnegative weights, w_g, that sum to 1, reflect the relative significance of each gage site with respect to site s. Their values

Fig. 11.1   River basin gage sites where streamflows are measured and recorded

will be based on the judgments of those who are familiar with the basin's hydrology.

$$Q_t^s = \left\{ \sum_g w_g Q_t^g / A^g \right\} A^s \qquad (11.1)$$

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 Eq. 11.1 to estimate streamflows will result in positive incremental flows. The downstream flow will be greater than the upstream flow. In arid regions runoff is not constant over the region. 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 less rainfall. In such arid areas the runoff is often less than the evapotranspiration and infiltration into the ground along the stream channel.

For stream channels where there exist relatively uniform conditions affecting loss loss, 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 gage flow, Q_t^g, and a loss rate per unit length of the stream or river, L^{gs} between gage site g and an ungaged site s. If there are at least two gage 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 may not be constant over the entire length between the gage stations, or even for all flow rates, however. Losses will likely 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 runoff, other methods may be required for estimating average streamflows at ungaged sites. The selection of the most appropriate method to use, as well as the most appropriate gage sites to use for estimating the streamflow, Q_t^s, at a particular site s can be a matter of judgment. The best gage site need not necessarily be the nearest gage to site s, but rather the site most similar with respect to important hydrologic variables.

The natural incremental flow between any two sites is simply the difference between their respective natural flows.

11.4 Streamflow Routing

If the duration of a within-year period is less than the time of flow throughout the stream or river system being modeled, and the flows vary within the system, 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 using 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 modeled, or if bank storage is not that significant, 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, that may take

into account losses, 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 (Chap. 5) using 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 preselected higher amount, the outflow can be 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 modeled 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:

$$O_{st} = \alpha\, S_{st} + \beta I_{st} \qquad (11.2)$$

Equation 11.2 applies for all time periods t and for all reach segments s in a particular reach. Different reaches will likely have different values of the parameters n, α, and β. The calibrated values of α and β are nonnegative 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 nonlinear function of the initial storage volume and inflow

$$O_{st} = (\alpha\, S_{st} + \beta I_{st})^{\gamma} \qquad (11.3)$$

The parameter γ is greater than 0 and no greater than 1. In practice γ is very close to 1. Again the values of these parameters, including the number of reaches n, can be found using nonlinear 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 by setting α and γ to 1, β 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 exist some river basin reaches where managers believe this particular routing applies.

11.5   Lakes and Reservoirs

Lakes and reservoirs are sites in a basin where surface water storage needs to be modeled. 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 modeled.

$$S_t + Q_t - R_t - L_t = S_{t+1} \qquad (11.4)$$

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, as discussed in Chaps. 4 and 8, 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.

$$S_t \leq K \quad \text{for each period } t. \qquad (11.5)$$

Equations 11.4 and 11.5 are the two fundamental equations required when modeling 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 together with 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 flood damage reduction, 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 (Chap. 9).

Reservoir storage capacity can be divided among the three major uses: (1) the *active storage* used for downstream flow regulation and for water supply, recreational development or hydropower production; (2) the the *dead storage* used for sediment collection; and (3) the *flood storage* capacity reserved to reduce potential downstream flood damage during flood events. These separate storage capacities are illustrated in Fig. 11.2. The distribution of active and flood control storage capacities may change over the year. For example there is no need for flood control storage in seasons that are not likely to experience floods.
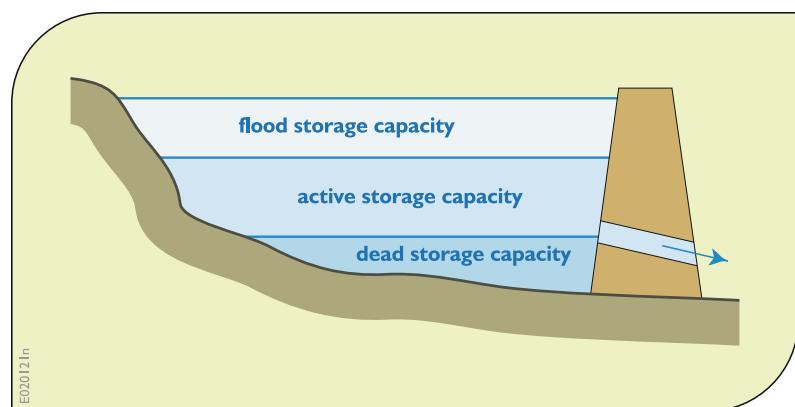
The next several sections of this chapter address how these capacities may be determined.

11.5.1   Estimating Active Storage Capacity

11.5.1.1   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 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 prespecified (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.

Fig. 11.2 Total reservoir storage volume capacity consisting of the sum of dead storage, active storage, and flood control storage capacities

flood storage capacity

active storage capacity

dead storage capacity

E020121n

Doing this over two cycles of the record of inflows will identify the maximum deficit volume associated with that release. 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.

Equation 11.6 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.
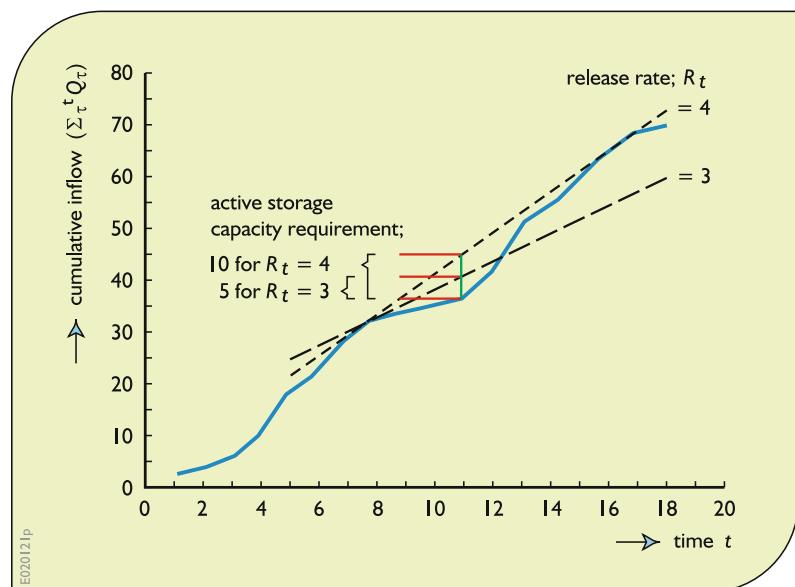
$$K_a = \text{maximum} \left[\sum_{t=i}^{j} (R_t - Q_t) \right], \quad (11.6)$$
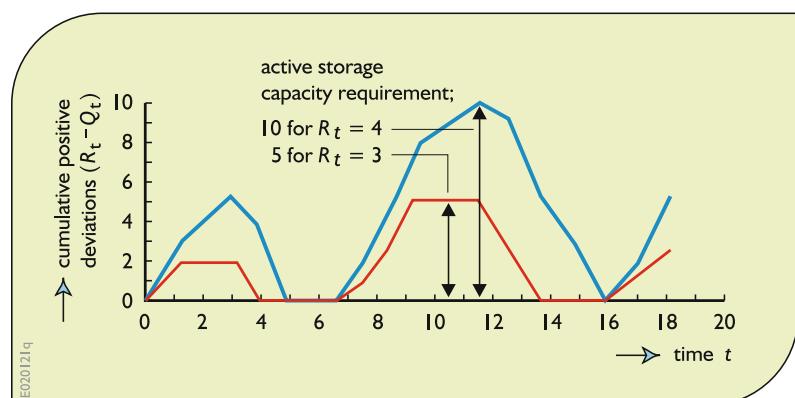
where $1 \leq i \leq j \leq 2T$.

Equation 11.6 is the analytical equivalent of graphical procedures proposed by Rippl for finding the active storage requirements. Two of these graphical procedures are illustrated in Figs. 11.3 and 11.4 for a 9-period inflow record of 1, 3, 3, 5, 8, 6, 7, 2, and 1. Rippl's original

Fig. 11.3 The Rippl or mass diagram method for identifying reservoir active storage capacity requirements. The releases R_t are assumed constant for each period t

Fig. 11.4 Alternative plot for identifying reservoir active storage capacity requirements

approach, shown in Fig. 11.3, involves plotting the cumulative inflow $\sum_{t=1}^{t} Q_\tau$ versus time t. Assuming a constant reservoir release, R, in each period t, a line with slope R is placed so that it is tangent to the cumulative inflow curve. To the right of these points of tangency the release R exceeds the inflow Q_t. The maximum vertical distance between the cumulative inflow curve and the release line of slope R equals the maximum water deficit, and hence the required active storage capacity. Clearly, if the average release R 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 nonnegative deviations, $\sum_{\tau}^{t} (R_\tau - Q_\tau)$, and note the biggest total deviation, as shown in Fig. 11.4.

These graphical approaches do not account for losses. Furthermore, the method shown in Fig. 11.3 is awkward if the desired releases in each period t are not the same. The equivalent method shown in Fig. 11.4 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 hydrologic records of flows are needed to identify the largest cumulative deficit inflow. After that the procedure will produce repetitive results. It is much easier to consider changing release values when determining the maximum deficit by the sequent peak method.

11.5.1.2 Sequent peak analyses

The sequent peak procedure is illustrated in Table 11.1. Let K_t be the maximum total storage requirement needed 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 Eq. 11.7 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 9-period record of 1, 3, 3, 5, 8, 6, 7, 2, and 1.

$$\begin{aligned} K_t &= R_t - Q_t + K_{t-1} \quad \text{if positive,} \\ &= 0 \qquad\qquad\quad \text{otherwise} \end{aligned} \tag{11.7}$$

The maximum of all K_t is the required storage capacity for the specified releases R_t and inflows, Q_t. Table 11.1 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 9 streamflows. Note this method does not require all releases to be the same.

11.5.2 Reservoir Storage-Yield Functions

Reservoir storage-yield functions define the minimum active active storage capacity required to insure a given constant release rate for a specified sequence of reservoir inflows. Mass diagrams, sequent peak analyses, and linear optimization (Chap. 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, 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

$$\text{Minimize } K_a \tag{11.8}$$

Table 11.1 Illustration of the sequent peak procedure for computing active storage requirements

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	

subject to

mass balance constraints

$$S_t + Q_t - R_t = S_{t+1} \quad t = 1, 2, \ldots, T; \quad T + 1 = 1 \tag{11.9}$$

capacity constraints

$$S_t \leq K_a \quad t = 1.2, \ldots, T \tag{11.10}$$

minimum release constraints

$$R_t \geq Y \quad t = 1.2, \ldots, T \tag{11.11}$$

for various values of the yield, Y.

Alternatively one can maximize the constant release yield, Y, for various values of active storage capacity, K_a, subject to the same constraint Eqs. 11.9–11.11.

$$\text{Maximize } Y \tag{11.12}$$
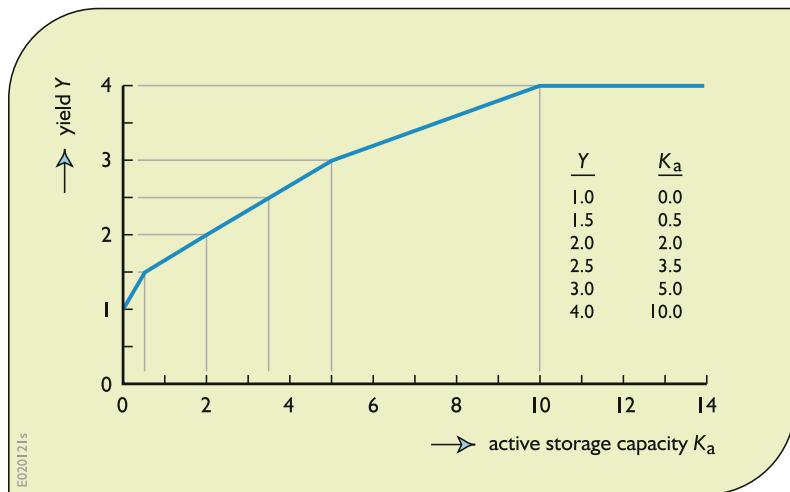
Constraints 11.9 and 11.11 can be combined to reduce the model size by T constraints.

$$S_t + Q_t - Y \geq S_{t+1} \quad t = 1, 2, \ldots, T; \quad T + 1 = 1 \tag{11.13}$$

The solutions of these two linear programming models, using the 9-period flow sequence referred

Fig. 11.5 Storage-yield
for the sequence of flows 1,
3, 3, 5, 8, 6, 7, 2, and 1

Y	K_a
1.0	0.0
1.5	0.5
2.0	2.0
2.5	3.5
3.0	5.0
4.0	10.0

to above and solved for various values of yield or capacity, respectively, are plotted in Fig. 11.5. 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 9 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 9 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 sstorage-yield function can be obtained by synthetic flows to supplement any measured streamflow record, taking parameter uncertainty

into account (as discussed in Chaps. 6 and 8). This will provide alternative sequences as well as more intervals between ranked flows.
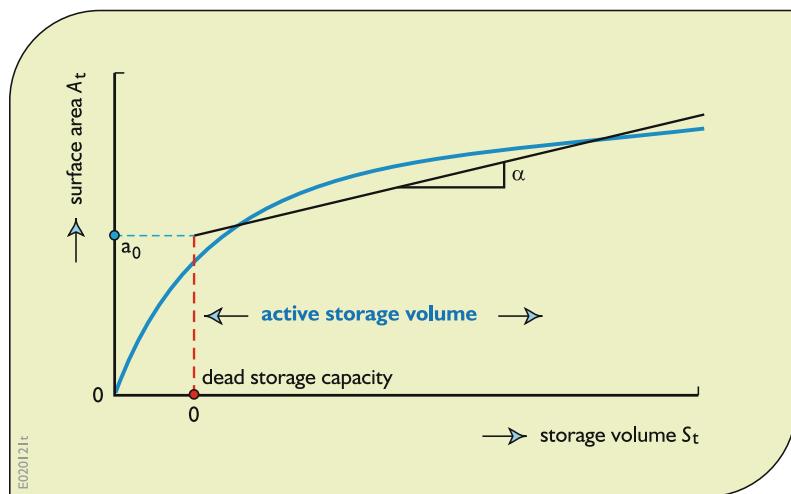
While the mass diagram and sequent peak procedures are relatively simple, they are not readily adaptable to reservoirs where evaporation losses and/or lake level regulation are important considerations, or to problems involving more than one reservoir. Mathematical programming (optimization) methods 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.

11.5.3   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 Fig. 11.6.

In addition to the storage area-volume function, seasonal surface water evaporation loss depths, E_t^{max}, must be assumed, perhaps based on measured evaporation losses over time.

Fig. 11.6 Storage surface area as a function of reservoir storage volume along with its linear approximation. The slope parameter a is the assumed increase in surface area associated with a unit increase in the storage volume

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

$$L_t = [a_o + a(S_t + S_{t+1})/2]E_t^{max} \qquad (11.14)$$

Letting

$$a_t = 0.5 \, a \, E_t^{max} \qquad (11.15)$$

the mass balance equation for storage volumes that include evaporation losses in each period t can be approximated as

$$(1 - a_t)S_t + Q_t - R_t - a_o E_t^{max} = (1 + a_t)S_{t+1} \qquad (11.16)$$

If Eq. 11.16 are used in optimization models for identifying preliminary designsof a proposed reservoir and if the active storage storage capacity turns out to be essentially zero, or just that required to provide for the fixed evaporation loss, $a_o E_t^{max}$, 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 simpler than introducing 0,1 integer variables that will remove the terms $a_o E_t^{max}$ in Eq. 11.16 if the active storage volume is 0 (using methods discussed in Chap. 4).

An alternative way to estimate evaporation loss that does not require a surface area—storage volume relationship, such as shown in Fig. 11.6, 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 Eqs. 11.9 or 11.13. This procedure assumes that 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.

11.5.4 Over- and Within-Year Reservoir Storage and Yields

An alternative approach to modeling 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 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 discusses a modeling approach for esti-mating over-year and within-year active storage requirements to deliver a specified annual and within-year period yields having a specified re-liability. The third and fourth parts expand this modeling 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 consisting 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 models having more within-year periods. Table 11.2 lists the nine years of

available streamflow data for each within-year season at a potential reservoir site. These streamflows are used to solve and compare the solutions of various yield models as well as to illustrate the concept of yield reliability.

11.5.4.1   Reliability of Annual yields

The maximum flow that can be made available at a specific site by the regulation of the historic streamflows from a reservoir of a given size is often referred to as the "firm yield" or "safe yield." These terms imply that the firm or safe yield is that yield which 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 may be better

Table 11.2 Recorded unregulated historical streamflows at a reservoir site

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

called an "historical yield." Historical and firm yield are often used synonymously.

A minimum 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, i.e., a probability, that it might not always be available in the future. There are some ways of estimating this probability.

Referring to the nine-year streamflow record listed in Table 11.2, 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 is greater than or equal to this value. In other words, it is the probability that this flow will be equaled 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 9 flows from the lowest to the highest value, and that the lowest value has a rank of 9.

Ranking the n flows of record from the highest to the lowest and assigning the rank m of 1 to the highest flow, and n to the lowest flow, the expected probability p that any flow of rank m will be equaled 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 equaled 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 $1/(1 − p)$.

11.5.4.2 Estimation of Active Reservoir Storage Capacities for Specified Yields

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 Eq. 11.7 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 11.1. The same yields can be obtained from a linear optimization model that minimizes active over-year storage capacity, K_a^o,

$$\text{Minimize } K_a^o \qquad (11.17)$$

This active over-year storage capacity must 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:

$$S_y + Q_y - Y_p - R_y = S_{y+1} \qquad (11.18)$$

$$S_y \leq K_a^o \qquad (11.19)$$

Once again, if the year index $y = n$, 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 11.2.

The over-year model, Eqs. 11.17–11.19, 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, Eq. 11.7, 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

$$\text{minimize } K_a \qquad (11.20)$$

subject to

$$S_{ty} + Q_{ty} - y_{pt} - R_{ty} = S_{t+1,y} \quad \forall t, y \quad (11.21)$$

$$S_{ty} \leq K_a \quad \forall t, y \qquad (11.22)$$

In Eq. 11.21, 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 within-year storage requirement, K_a^w, is the difference in the active capacities resulting from these two models, Eqs. 11.17–11.19, and Eqs. 11.20–11.22.

Table 11.3 shows some results from solving both of the above models. The over-year storage capacity requirements, K_a^o, are obtained from Eqs. 11.17–11.19. The combined over-year and within-year capacities, K_a, are obtained from solving Eqs. 11.20–11.22. The difference between the over-year storage capacity, K_a^o, 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 (Eqs. 11.20–11.22) 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.
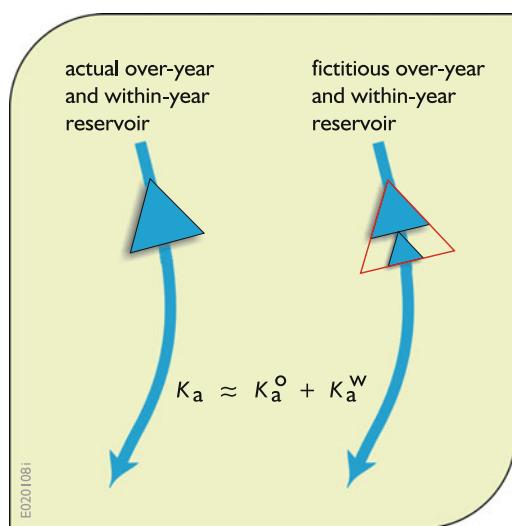
Even though the severity of future droughts is unknown, many planners accept the traditional practice of using the historical critical drought period for reservoir design and operation studies on the assumption that having observed such an event in the past, it is certainly possible to experience similar conditions 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 (Chap. 6) are more acceptable as a basis for yield estimation.

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 (Fig. 11.7). The upper reservoir is the over-year storage reservoir, whose capacity required for an annual yield is determined by an over-year model, e.g., Eqs. 11.17–11.19. The purpose of the "downstream" within-year reservoir is to distribute as desired in each within-year period t a portion of the annual yield produced by the "upstream" over-year reservoir. Within-year storage capacity would not be needed if the distribution of the average inflows into the upper over-year reservoir exactly coincided with the desired distribution of within-year yields. Otherwise 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.

Table 11.3 Active Storagerequirements for various within-year yields

annual yield	within-year yields		required active storage volume capacity		
			within -year	over -year	total
$y\ 0.9$	$t = 1$	$t = 2$	K_a^w	K_a^o	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

Fig. 11.7 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

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 β_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.

$$\text{Minimize } K_a^w \qquad (11.23)$$

subject to

$$s_t + \beta_t Y_p - y_{pt} = s_{t+1} \quad \forall t \quad T+1 = 1 \qquad (11.24)$$

$$s_t \leq K_a^w \quad \forall t \qquad (11.25)$$

Since the sum of β_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_p.

The over-year model, Eqs. 11.17–11.19, and within-year model, Eqs. 11.23–11.25, can be combined into a single model for an n-year sequence of flows

$$\text{Minimize } K_a \qquad (11.26)$$

subject to

$$S_y + Q_y - Y_p - R_y = S_{y+1} \quad \forall y$$
$$\text{if} \quad y = n, \quad y + 1 = 1 \qquad (11.27)$$

$$S_y \leq K_a^o \quad \forall y \qquad (11.28)$$

$$s_t + \beta_t Y_p - y_{pt} = s_{t+1} \quad \forall t$$
$$\text{if} \quad t = T, \quad T + 1 = 1 \qquad (11.29)$$

$$s_t \leq K_a^w \quad \forall t \qquad (11.30)$$

$$\sum_t y_{pt} = Y_p \qquad (11.31)$$

$$K_a \geq K_a^o + K_a^w \qquad (11.32)$$

Constraint 11.31 is not required due to Eq. 11.29, but is included here to make it clear that the sum of within-year yields will equal the over-year yield. 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 Eq. 11.30 can be combined with Eq. 11.32, saving a constraint. If this is done, the combined model contains $2n + 2T + 1$ constraints, compared to the more accurate model, Eqs. 11.20–11.22, that contains $2nT$ constraints.

If the fractions β_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 11.2, 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 Tables 11.2 or 11.3 we can see that 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 equaling 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 11.3.

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 as the desired annual yield since it is the upper limit of the annual yield. For example in Table 11.2 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

$$S_7 + Q_7 + Q_8 + Q_9 + Q_1 - 4Y_p - R_7 = S_2$$
$$(11.33)$$

This saves a total of 3 over-year continuity constraints and 3 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 Eq. 11.28 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 modeled 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 9-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 can be omitted from Eq. 11.27 in that 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 0.90 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 Eq. 11.27, and so on.

The number of years of yield failure determines the estimated reliability of each yield. An annual yield that fails in f years has an estimated probability $(n - f)/(n + 1)$ of being equaled or exceeded in any future year. Once the desired reliability of a yield is known, the problem is to select the appropriate failure years and to specify the permissible extent of failure in those f failure years.

To consider different yield reliabilities p let the parameter α_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 α_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 α_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.

$$S_y + Q_y - \alpha_y^p Y_p - R_y = S_{y+1} \quad \forall y \quad \text{if} \atop y = n, \quad y + 1 = 1 \tag{11.34}$$

When writing Eq. 11.34, 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 7-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 7 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 2–4. 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 low-flow 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 9-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 andwithin-year continuity constraints can be written

$$S_y + Q_y - Y_{0.9} - \alpha_y^{0.7} Y_{0.7} - R_y = S_{y+1} \quad \forall y$$
$$\text{if} \quad y = n, \quad y + 1 = 1$$
$$(11.35)$$

$$s_t + \beta_t(Y_{0.9} + Y_{0.7}) - y_{0.9,t} - y_{0.7,t} = s_{t+1} \quad \forall t$$
$$\text{if} \quad t = T, \quad T + 1 = 1$$
$$(11.36)$$

Now an additional constraint is needed to insure that each within-year yield of a reliability p adds up to the annual yield of the same reliability. Selecting the 90% reliable yield,

$$\sum_t y_{0.9,t} = Y_{0.9} \qquad (11.37)$$

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 Eq. 11.14 (see also Fig. 11.6) results in an estimated evaporation loss L_{yt}.

$$L_{yt} = \left[a_o + a(S_y + (s_t + s_{t+1})/2)\right] E_t^{\max}$$
$$(11.38)$$

Summing L_{yt} over all within-year periods t defines the estimated annual evaporation loss, E_y.

$$E_y = \sum_t \left[a_o + a(S_y + (s_t + s_{t+1})/2)\right] E_t^{\max}$$
$$(11.39)$$

This annual evaporation loss applies, of course, only when there is a nonzero active storage capacity requirement. These annual evaporation losses can be included in the over-year continuity constraints, such as Eq. 11.35. If they are, the assumption is being made that their within-year distribution will be defined by the fractions β_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 Eq. 11.38 over all years y divided by the total number of years, n.

$$L_t = \sum_y^n \left[a_o + a(S_y + (s_t + s_{t+1})/2)\right] E_t^{\max}/n$$
$$(11.40)$$

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 Eq. 11.37 and 11.38, for p equal to 0.9 in the example,

$$Y_p = \sum_t \left\{ y_{pt} + \sum_y^n \left[a_o + a(S_y + (s_{t+1})/2)\right] E_t^{\max}/n \right\}$$
$$(11.41)$$

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 aid in defining a reservoir operating policy for such simulation studies.

Reservoir Operation Rules

Reservoir operation rules 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 volumes 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.
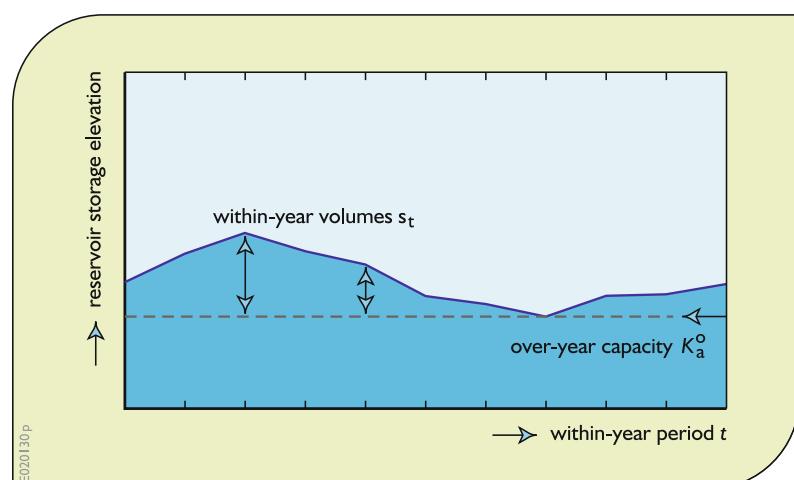
To construct an operation rule that identifies storage zones, each having a specific release policy, the values of the dead and flood storage capacities, K_D and K_f are needed together with the over-year storage capacity, K_a^o, and within-year storage volumes s_t in each period t. Since both K_a^o 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 9-year flow record and 10 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 Fig. 11.8.

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 this shaded zone shown in Fig. 11.8, 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.

Fig. 11.8 Reservoir release rule showing the identification of the most reliable release zone associated with the within-year yields $y_{0.9,t}$

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 Fig. 11.9.
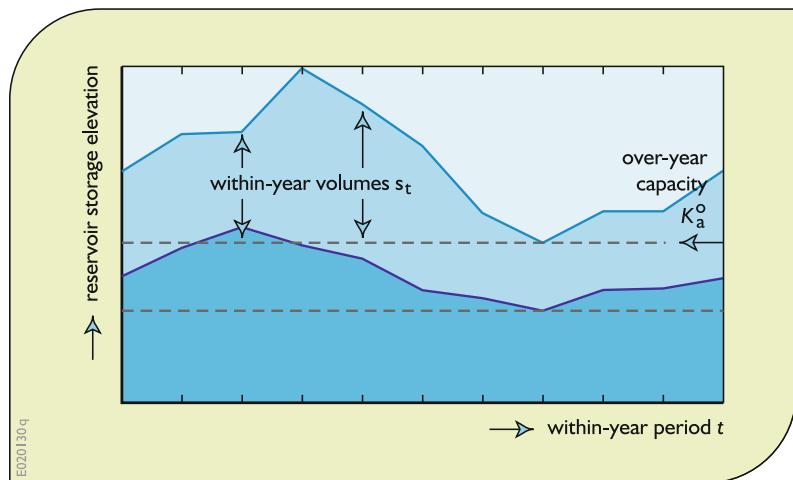
If at any time t the actual storage volume is in the second lighter shaded zone in Fig. 11.9, both the release should be the sum of the most reliable yield, $y_{0.9,t}$ and the incremental secondary yield $y_{0.7,t}$. If only these releases are made, the probability of being in that zone, when simulated, 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 Fig. 11.23, then a release

can be made to satisfy any downstream demand. Converting storage volume to elevation, this release policy is summarized in Fig. 11.10.
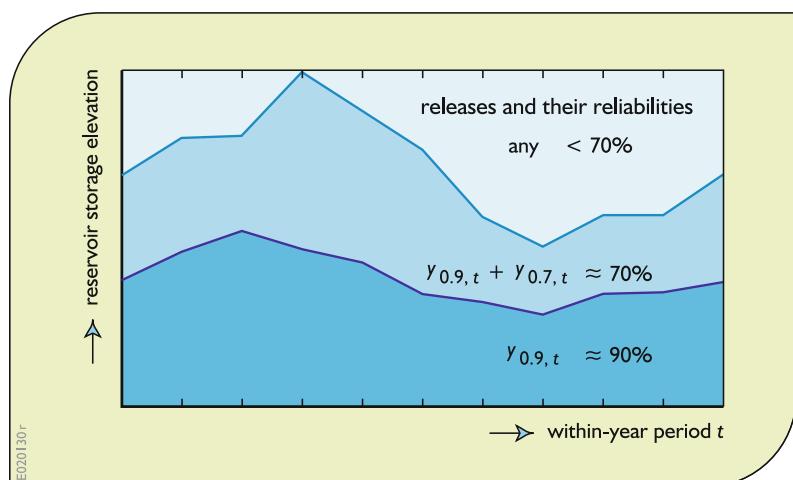
These yield models focus only on the active storage capacity requirements. They can be a part of a model that includes flood storage requirements as well (as previously discussed in this chapter). If the actual storage volume is within the flood control zone in the flood season, releases should be made 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 these rules

Fig. 11.9 Reservoir release rule showing the identification of the second most reliable release zone associated with the total within-year yields $y_{0.9,t} + y_{0.7,t}$

Fig. 11.10 Reservoir release rule defined by the yield model

should be simulated, evaluated, and refined prior to their actual adoption.

11.6 Drought and Flood Risk Reduction

11.6.1 Drought Planning and Management

Droughts are natural hazards that unlike floods, tornadoes, and hurricanes, occur slowly and gradually over a period of time. The absence of a precise drought threshold introduces some uncertainty about whether a drought exists and, if it does, its degree of severity. The impacts of drought are nonstructural and typically spread over a larger geographical area than are damages resulting from other natural hazards. All of these drought characteristics have impacted the development of effective drought preparedness plans.

Droughts result from a deficiency of precipitation compared to normal (long-term average) amounts that, when extended over a season or especially over a longer period of time, is insufficient to meet the demands of human activities and the environment. All types of drought results in water shortages for one or more water-using activities.

Droughts differ from one another in three essential characteristics: intensity, duration, and spatial coverage. Moreover, many disciplinary perspectives of drought exist. Because of these numerous and diverse disciplinary views, confusion often exists over exactly what constitutes a drought. Regardless of such disparate views, the overriding feature of drought is its negative impacts on people and the environment.

11.6.1.1 Drought Types

Droughts are normally distinguished by type: meteorological, hydrological, agricultural, and socioeconomic. Meteorological drought is expressed solely on the basis of the degree of dryness in comparison to some normal or average amount and the duration of the dry period. Drought intensity and duration are the key descriptors of this type of drought. Agricultural

drought links various characteristics of meteorological drought to agricultural impacts, focusing on precipitation shortages, differences between actual and potential evapotranspiration, and soil water deficits.

Hydrological droughts are described based on the effects of low precipitation on surface or subsurface water supply (e.g., streamflow, reservoir storage, lake levels, and groundwater) rather than with precipitation shortfalls. Hydrological droughts usually lag the occurrence of meteorological and agricultural droughts because more time elapses before precipitation deficiencies are detected in rivers, reservoirs, groundwater aquifers, and other components of the hydrologic system. As a result, hydrological droughts are typically detected later than other drought types. Water uses affected by drought can include multiple purposes such as power generation, flood control, irrigation, domestic drinking water, industry, recreation, and ecosystem preservation.

Socioeconomic droughts are linked directly to the supply of some economic good. Increases in population can alter substantially the demand for these economic goods over time. The incidence of socioeconomic drought can increase because of a change in the frequency of meteorological drought, a change in societal vulnerability to water shortages, or both. For example, poor land use practices such as overgrazing can decrease animal carrying capacity and increase soil erosion, which exacerbates the impacts of, and vulnerability to, future droughts.

11.6.1.2 Drought Impacts

The impacts of drought are often widespread through the economy. They can be direct and indirect. Restrictions in water use resulting from drought is a direct or first-order impact of drought. However, the consequences of such restrictions could result in loss of income, farm and business foreclosures, and government relief programs) are possible indirect second- or third-order impacts.

The impacts of drought appear to be increasing in both developing and developed countries, which in many cases reflects the persistence of

non-sustainable development and population growth. Lessening the impacts of future drought events typically requires the development of drought risk policies that emphasize a wide range of water conservation and early warning measures. Drought management techniques are often conditional on the severity of the drought. Identifying the actions to take and the thresholds indicating when to take them are best accomplished prior to a drought, as agreements among stakeholders are easier to obtain when individuals are not having to deal with the impacts of an ongoing drought.

Drought impacts can be economic, environmental, and social.

Economic impacts can include direct losses to agricultural and industrial users, losses in recreation, transportation, and energy sectors. Other indirect economic impacts can include resulting unemployment and loss of tax revenue to local, state, and federal governments.

Environmental losses include damages to plant and animal species in natural habitats, and reduced air and water quality; an increase in forest and range fires; the degradation of landscape quality; and possible soil erosion. These losses are difficult to quantify, but growing public awareness and concern for environmental quality has forced public officials to focus greater attention on them.

Social impacts can involve public safety, health, conflicts among water users, and inequities in the distribution of impacts and disaster relief programs. As with all natural hazards, the economic impacts of drought are highly variable within and among economic sectors and geographic regions, producing a complex assortment of winners and losers with the occurrence of each disaster.

11.6.1.3   Drought Preparedness and Mitigation

Droughts happen, and it makes no sense to wait until realizing a drought is happening before preparing plans and policies to mitigate the adverse impacts from a drought. As evidenced by the ongoing drought (at this writing) in California, and the even more severe drought those in southeastern Australia recently witnessed, drought management has to involve the institutions that not only manage water supply systems, but all those who use water, and all those who make land-use decisions that impact water runoff. It can involve hydrologic modeling methods discussed in Chap. 6, and reservoir modeling as discussed in Chaps. 4 and 8. Appendix C of this book (contained on a disk or downloadable from the web) discusses drought management modeling methods and options in more detail.

11.6.2   Flood Protection and Damage Reduction

Next consider the other extreme—floods. Two types of structural alternatives are often used for flood risk reduction. One is the provision of flood storage capacity in reservoirs designed to reduce downstream peak flood flows. The other is channel enhancement and/or flood-proofing structures that are designed to contain peak flood flows and reduce damage. This section introduces methods of modeling 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 modeling 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.

11.6.2.1   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 flood season of the year, as shown in Fig. 11.2. 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 that typically include time periods of a week, or 10 days, or months or seasons spanning several months. If we modeled these short daily or hourly 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 variables within longer period optimization models .

Consider a potential flood damage site along a river. A flood control reservoir can be built upstream of that potential damage site. The question is how much flood storage capacity, if any, should the reservoir contain. For various assumed capacities and operating policies, 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, water elevations and associated economic flood damages on the floodplain can be estimated. To calculate the expected annual damages 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 described by its expected return period. How many years would one expect to wait, on average, to observe another flood of equal or greater than a flood of some specified magnitude? This is the reciprocal of the probability of observing such a flood or greater in any given year. A T-year flood has a probability of being equaled 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 1000 or 10,000 years,

divided by 1000 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.
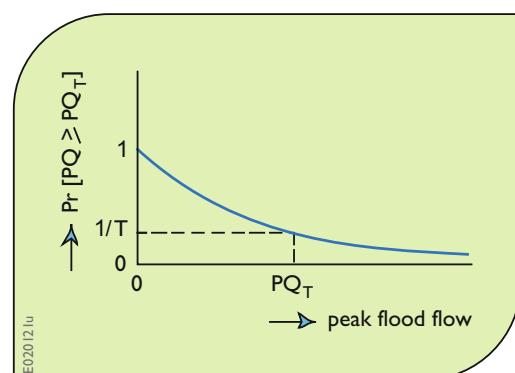
If PQ is the random annual peak flood flow and PQ_T is a particular peak flood flow having a return period of T years, then by definition the probability of an actual flood of PQ equaling or exceeding PQ_T is $1/T$.

$$\Pr[PQ \geq PQ_T = 1/T] \qquad (11.42)$$

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 11.42 is plotted in Fig. 11.11.

The exceedance probability distribution shown in Fig. 11.11 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 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

Fig. 11.11 Probability of annual peak flood flows being exceeded

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
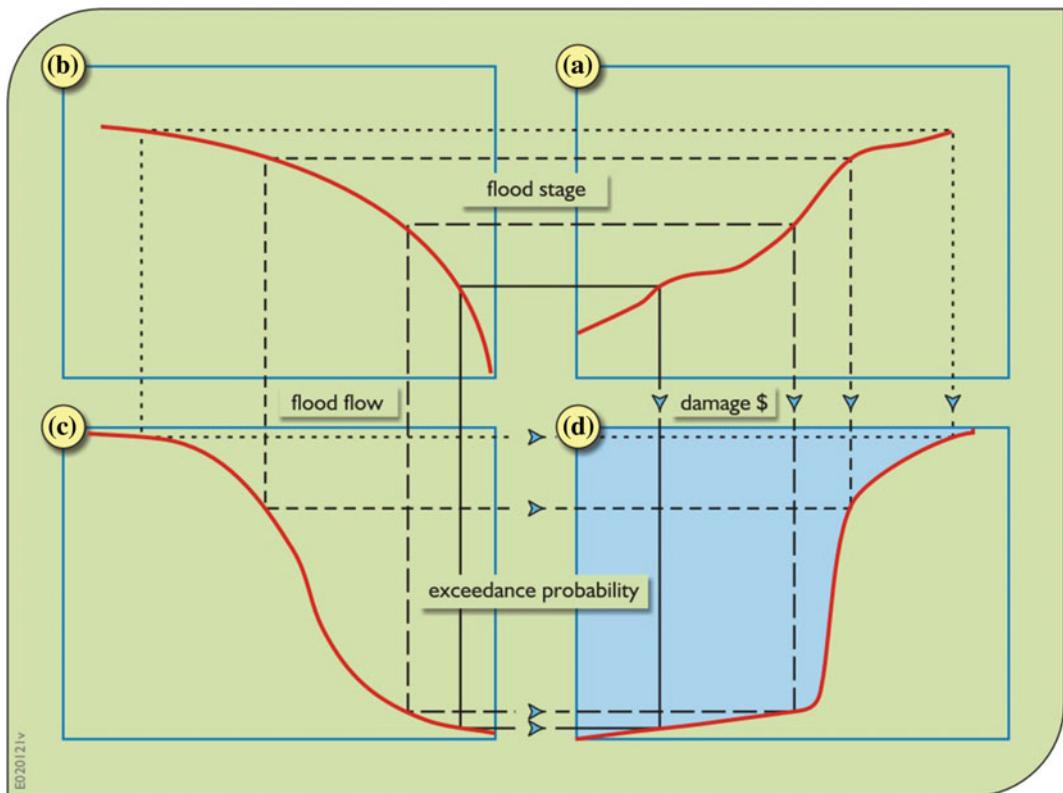
$$\Pr[FD \geq FD_T] = 1/T \qquad (11.43)$$

The area under this exceedance probability distribution is the expected annual flood damage , $E[FD]$.

$$E[FD] = \int_0^\infty \Pr[FD \geq FD_T]dFD_T \qquad (11.44)$$

This computational process is illustrated graphically in Fig. 11.12. 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 equaling 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 Fig. 11.12, must be known. These do not depend on the flood storage

Fig. 11.12 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**)
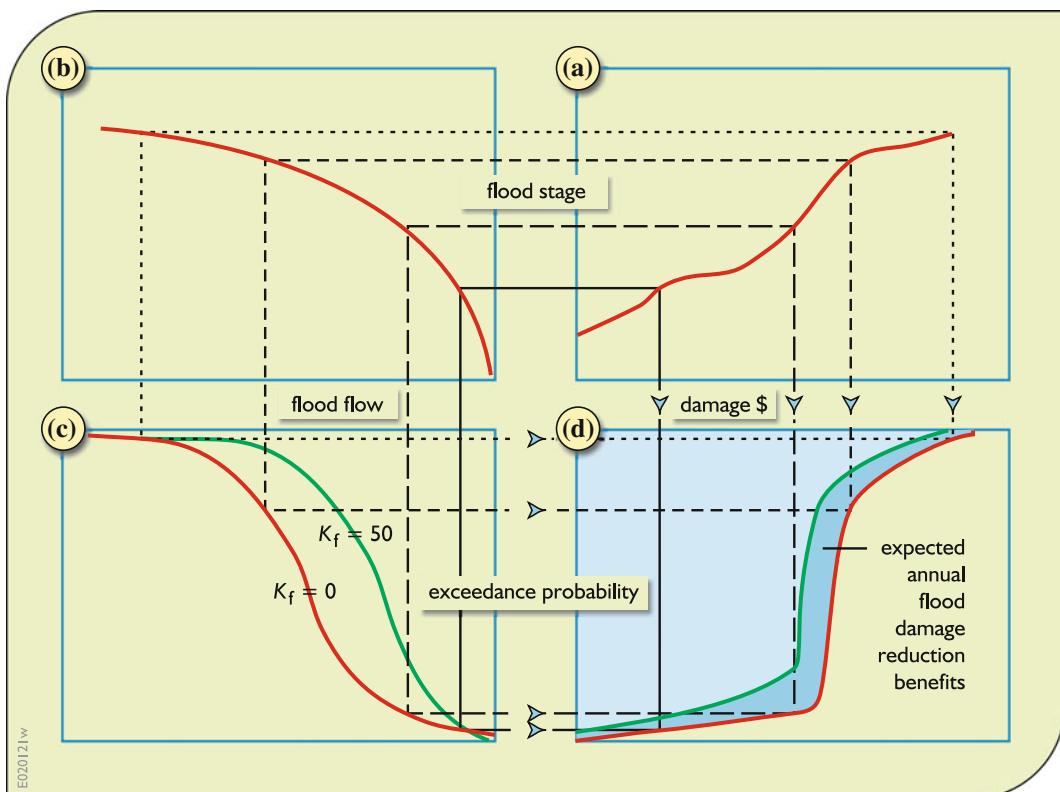
capacity in an upstream reservoir. The information in quadrant (c) is similar to that shown in Fig. 11.11 defining the exceedance probabilities of each peak flow. Unlike the other three functions, this distribution depends on the upstream flood storage capacity and flood flow release policy. This peak 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 Fig. 11.13. 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 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.

Fig. 11.13 Calculation of expected annual flood damage reduction benefits, shown as the darkened portion of quadrant (**d**), associated with a specified reservoir flood storagecapacity

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.

$$K \geq K_d + S_t + K_f \quad \text{for all periods } t \text{ in flood season}$$
$$\text{plus the following period that}$$
$$\text{represents the end of}$$
$$\text{the flood season}$$

$$(11.45)$$

$$K \geq K_d + S_t \quad \text{for all remaining periods } t$$
$$(11.46)$$

In the above equations the dead storage capacity, K_d, is assumed known. It is included in the capacity Eqs. 11.45 and 11.46 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.

11.6.2.2 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 dikes 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 prespecified 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.

$$QN_T \leq QS + QR \qquad (11.47)$$
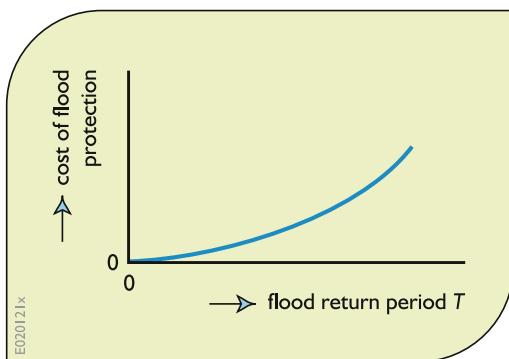
The extent to which a specified upstream reservoir flood storage capacity reduces the design peak flow at the downstream potential damage site can be obtained by routing the design flood through the reservoir and the channel between the reservoir and the downstream site. Doing this for a number of reservoir flood storage capacities permits the definition of a peak flow reduction function, $f_T(K_f)$.

$$QS = f_T(K_f) \qquad (11.48)$$

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:

$$\text{minimize } \text{Cost}_K(K_f) + \text{Cost}_R(QR) \qquad (11.49)$$

Fig. 11.14 Tradeoff between minimum cost of flood protection and flood risk, as expressed by the expected return period

subject to

$$QN_T \leq f_T(K_f) + QR \qquad (11.50)$$

Equations 11.49 and 11.50 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, i.e., how much flood storage capacity and how much levee protection.

Solving Eqs. 11.47 and 11.48 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 Fig. 11.14.

11.6.2.3 Estimating Risk of Levee Failures

Levees are built to reduce the likelihood of flooding on the flood plain. Flood flows prevented from flowing over a floodplain due to a levee will have relatively little effect on users of the flood plain, unless of course the levee fails to contain the flow. Levee failure can result from flood events that exceed (overtop) the design capacity of the levee. Failure can also result from various types of geostructural weaknesses. If any of the flow in the stream or river channel passes over, through or under the levee and onto the flood plain, the levee is said to fail. 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 is 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 damages further downstream.
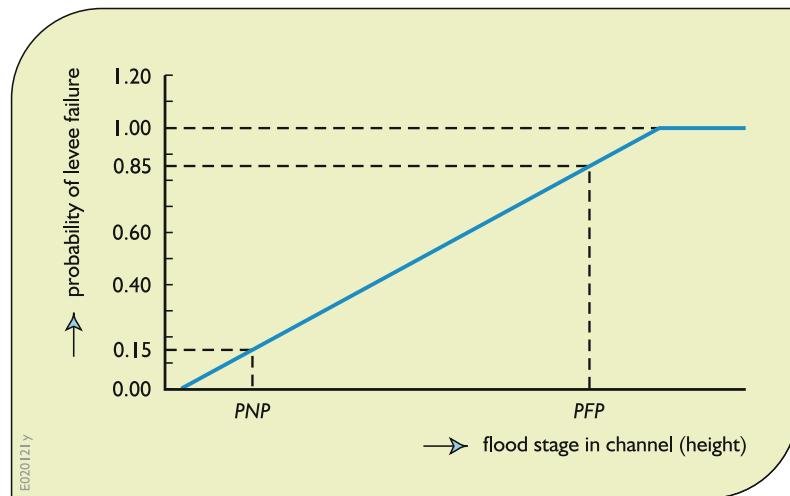
The probability of levee failure given the flood stage (height) in the stream or river channel is often modeled using two flood stages. The US Army Corps of Engineers calls the lower stage the probable non-failure point, *PNP*, and the higher stage is called the probable failure point, *PFP* (USACE 1991). 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 Fig. 11.15. 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.
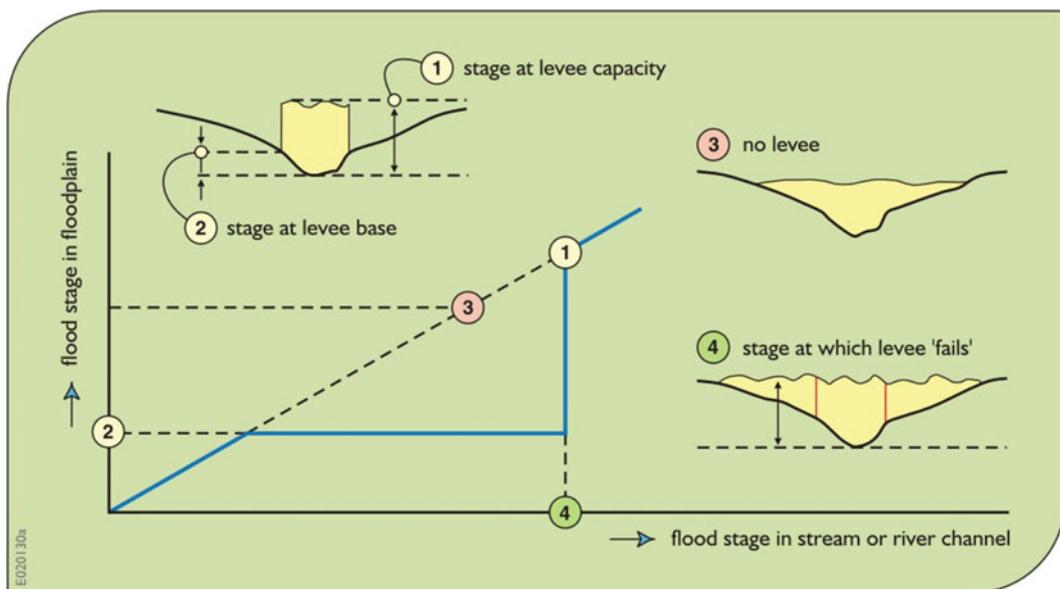
To estimate the risk of a flood in the floodplain protected by a levee due to overtopping or geo-structural levee failure, the relationships between flood flows and flood stages in the channel and on the floodplain must be defined, just as it had to be to carry out the analyses shown in Figs. 11.12 and 11.13.

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 flood plain. Figure 11.16 illustrates the relationship between the flood stage in the channel and the flood stage in the flood plain, assuming no geo-structural failure of the levee. Obviously once the flood begins overtopping the levee, the flood stage in the flood plain 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 11.17 illustrates the relationship between flood flow and flood stage in a

Fig. 11.15 Assumed cumulative probability distribution of levee failure expressed as function of flood stage in river channel

Fig. 11.16 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
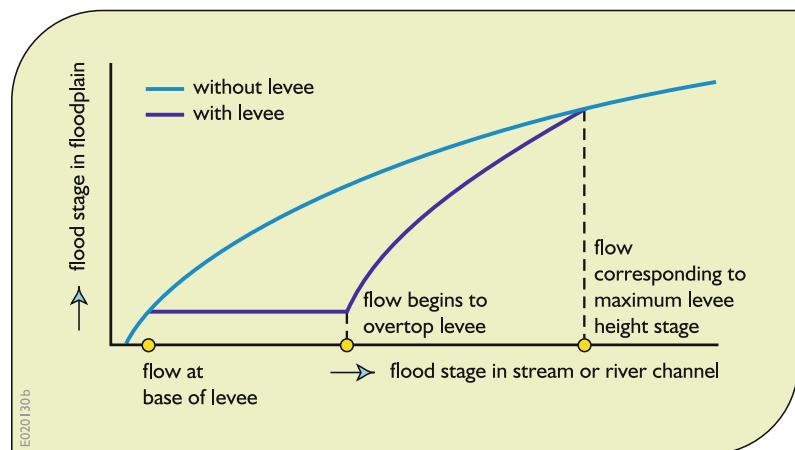
floodplain with and without flood levees, again assuming no geo-structural levee failure.
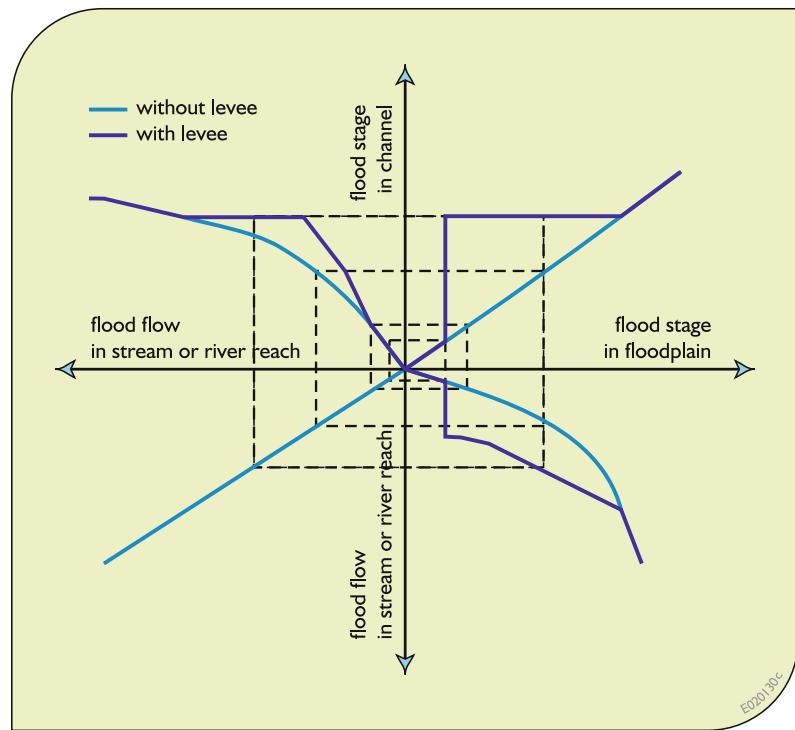
Combining Figs. 11.16 and 11.17 defines the relationship between reach flow and channel stage. This is illustrated in the upper left quadrant of Fig. 11.18.

Combining the relationship between flood flow and flood stage in the channel (upper left quadrant of Fig. 11.18) with the probability distribution of levee failure (Fig. 11.15) and the probability distribution of annual peak flows being equaled or exceeded (Fig. 11.11), provides

Fig. 11.17 Relationship between flood flow and flood stage in a floodplain with and without flood levees, again assuming no geo-structural levee failure

Fig. 11.18 Deriving the relation (shown in the *upper left quadrant*) between flood flow and flood stage in the channel
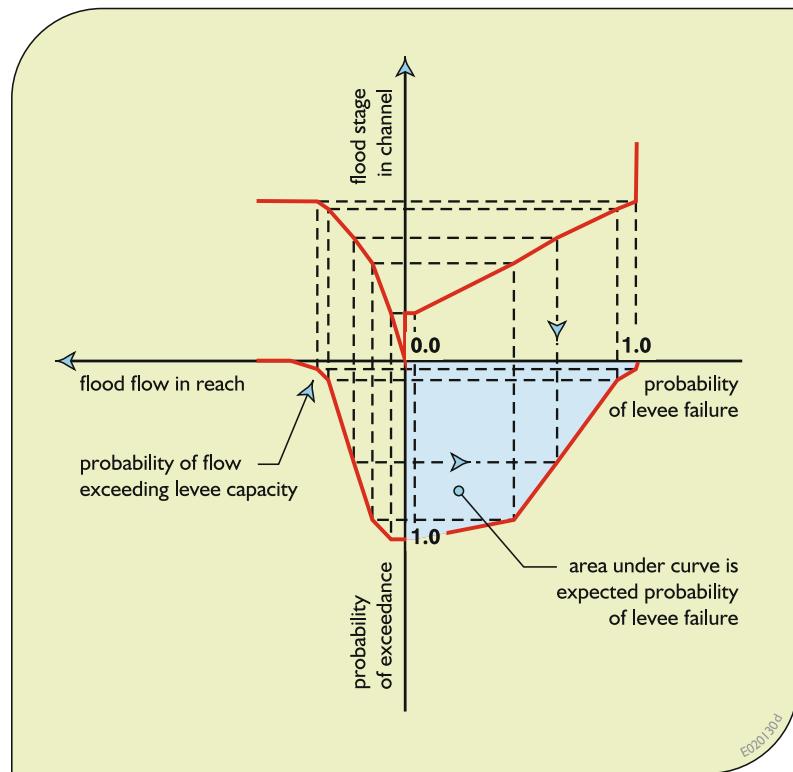
an estimate of the expected probability of levee failure. Figure 11.19 illustrates this process of finding, in the lower right quadrant, the shaded area that equals the expected annual probability of levee failure from overtopping and/or geo-structural failure.

The channel flood-stage function, $S(q)$, of peak flow q shown in the upper left quadrant of Fig. 11.19 is obtained from the upper left

quadrant of Fig. 11.18. The probability of levee failure, $PLF(S)$, a function of flood stage, $S(q)$, shown in the upper right quadrant is the same as in Fig. 11.15. The annual peak flow exceedance probability distribution, $F_Q(q)$, (or its inverse $Q(p)$) in the lower left quadrant is the same as Fig. 11.12 or that in the lower left quadrant (c) of Fig. 11.13. The exceedance probability function in the lower right quadrant of Fig. 11.19 is

Fig. 11.19 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

derived from each of the other three functions, as indicated by the arrows, in the same manner as described in Fig. 11.12.

In mathematical terms, the annual expected probability of levee failure, $E[\mathrm{PLF}]$, found in the lower right quadrant of Fig. 11.19, equals

$$
\begin{aligned}
E[\mathrm{PLF}] &= \int_0^\infty \mathrm{PLF}[S(q)]f(q)\mathrm{d}q \\
&= \int_0^1 \mathrm{PLF}[S(Q(p))]\mathrm{d}p = \int_0^1 \mathrm{PLF}'(p)\mathrm{d}p,
\end{aligned}
$$

(11.51)

where $\mathrm{PLF}'(p)$ 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
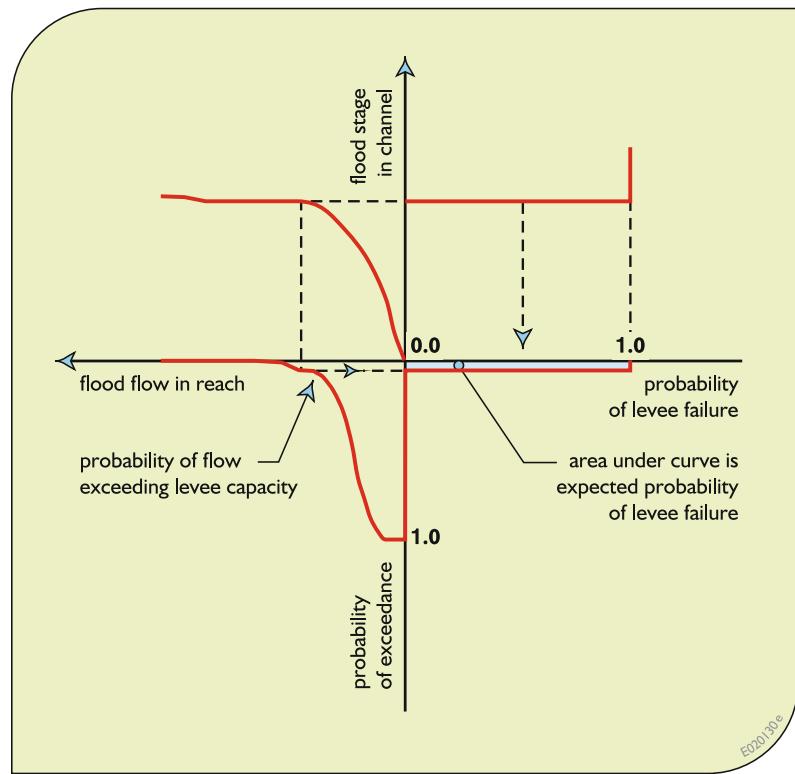
exceeding a channel flow whose stage equals the levee height, as defined in the lower left quadrant of Fig. 11.19. This is shown in Fig. 11.20.

Referring to Fig. 11.20, 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.

11.6.2.4   Annual Expected Damage from Levee Failure

A similar analysis can provide an estimate of the expected annual flood plain damage for a stream or river reach. Consider, for example, a parcel of land on a flood plain at some location i. If an economic efficiency objective were to guide the development and use of this parcel, the owner would want to maximize the net annual economic benefits derived from its use, B_i, less the

Fig. 11.20 Calculation of annual probability of equaling 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
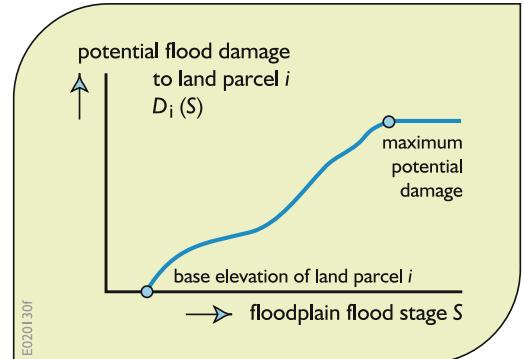
annual (non-flood damage) costs, C_i, and the expected annual flood damages, EAD_i. The issue of concern here is the estimation of these expected annual flood damages.

Damages 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, etc.). 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 Fig. 11.21.

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.

$$E[D_i] = \int D_i(S)F_s(S)\mathrm{d}S \qquad (11.52)$$

Fig. 11.21 An example function defining the damages that will occur given any flood stage S to a parcel of land i

The sum of these 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.

$$EAD = \sum_i E[D_i] \qquad (11.53)$$

Alternatively the annual expected flood damage could be based on a calculated probability of exceeding a specified flood damage, as shown in Fig. 11.12. 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.

$$D(S) = \sum_i D_i(S) \qquad (11.54)$$

This is the function shown in quadrant (a) in Fig. 11.12.

Levee failure probabilities, $PLF'(p)$, based on the exceedance probability p of peak flows, or stages, as defined in Fig. 11.31 and Eq. 11.49 can be included in calculations of expected annual damages. 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.

$$EAD = \int_0^1 D'(p) \, PLF'(p) \, dp \qquad (11.55)$$

Note that the flood plain damages and probability of levee failure functions in Eq. 11.55 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 Fig. 11.22. The "without levee" function in the lower right quadrant of Fig. 11.18, 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 Fig. 11.18, i.e., 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 Fig. 11.22.

11.6.2.5   Risk-Based Analyses
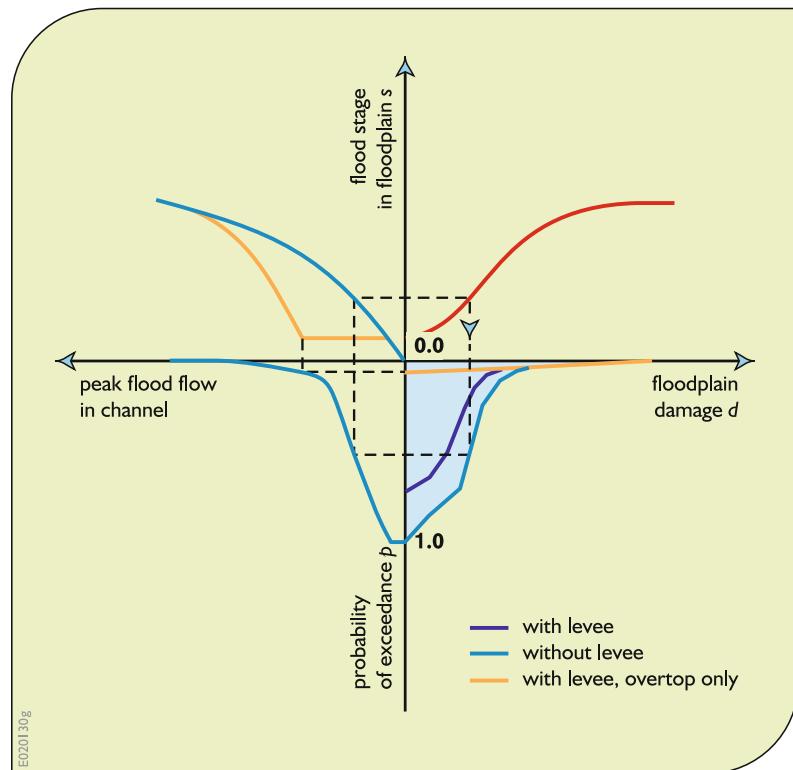
Risk-based analyses attempt to identify the uncertainty associated with each of the inputs used to define the appropriate capacities of various flood risk reduction measures. There are numerous sources of uncertainty associated with each of the functions shown in quadrants (a), (b), and (c) in Fig. 11.25. This uncertainty translates to uncertainty associated with estimates of flood risk probabilities and expected annual flood damage reductions obtained from reservoir flood storage capacities and channel improvements.

Going to the substantial effort and cost of quantifying these uncertainties, which themselves will be surely be uncertain, does however provide additional information. The design of any flood protection plan can be adjusted to reflect 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 Fig. 11.12) is defined from a risk-based-analysis. Figure 11.23 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 for an equivalent expected $T + \Delta$ year return period flow, $PQ_{T+\Delta}$, i.e., 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.
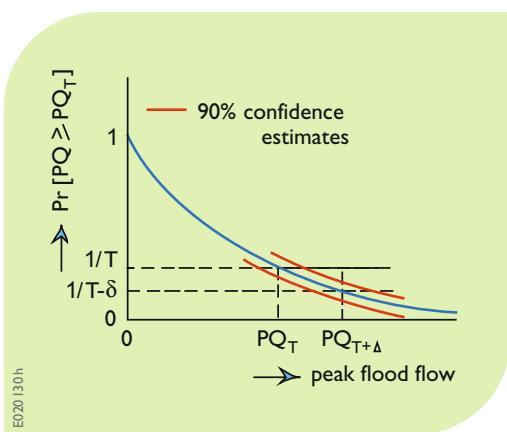
If society wanted to eliminate flood damage it could do it, but at a high cost. This would require either costly flood control structures or eliminating economic activities on lands subject to possible flooding. Both reduce expected economic returns from the floodplain. Hence such actions are not likely to be taken. There will

Fig. 11.22 Calculation of expected annual flood damage taking into account probability of levee failure

E020130g
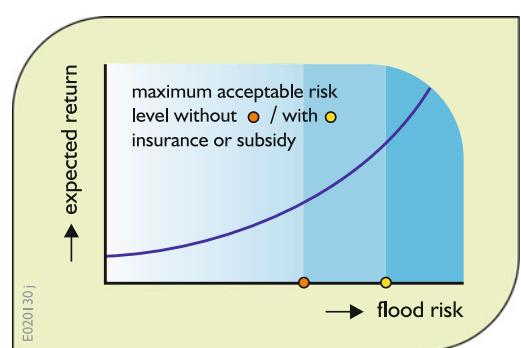
always be a risk of flood damage. Analyses such as those just presented help identify these risks. Risks can be reduced and managed but not

E020130h

Fig. 11.23 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 equaled or exceeded

eliminated. Finding the best levels of flood protection and flood risk, together with risk insurance or subsidies (illustrated in Fig. 11.24) is the challenge for public and private agencies alike. Floodplain management is as much concerned with good things not happening on them as with bad things—like floods—happening on them.

E020130j

Fig. 11.24 Relationship between expected economic return from flood plain 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
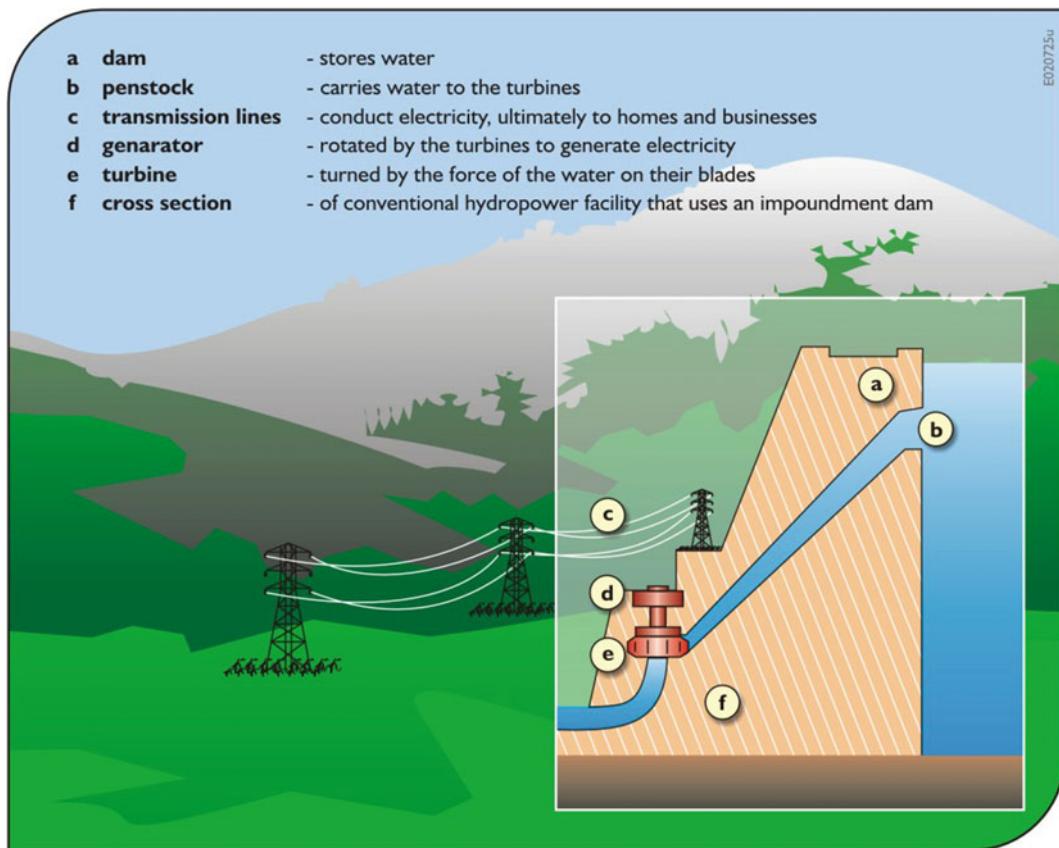
11.7   Hydroelectric Power Production

Hydropower plants, Fig. 11.25, convert the energy from the flowing water to mechanical and then electrical energy. These plants containing turbines and generators are typically located either in or adjacent to dams. Pipelines (penstocks) carry water under pressure from the reservoir to the powerhouse. Power transmission systems transport the produced electrical energy from the powerhouse to where it is needed.

The principal advantages of using hydropower are the absence of polluting emissions during operation, its capability to respond relatively quickly to changing utility load demands, and its relatively low operating costs. Disadvantages can include high initial capital cost and potential site-specific and cumulative environmental impacts. Potential environmental impacts of hydropower projects include altered flow regimes below storage reservoirs or within diverted stream reaches, water quality degradation, mortality of fish that pass through turbines, blockage of fish migration, and flooding of terrestrial ecosystems by impoundments. However, in many cases, proper design and operation of hydropower projects can mitigate some of these impacts. Hydroelectric projects can also provide additional benefits such as from recreation in reservoirs or in tailwaters below dams.

Hydropower plants can be either *conventional* or *pumped storage*. Conventional hydropower 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

Fig. 11.25 Hydropower system components

water resources may be used. Competing water uses may include irrigation, flood control, navigation, downstream flow dilution for quality improvement, and municipal and industrial water supply. Pumped storage plants pump the water, usually through a reversible turbine that acts as a pump, from a lower supply source to an upper reservoir. While pumped storage facilities are net energy consumers, they are income producers. They are valued by a utility because they can be brought online rapidly to operate in a peak power production mode when energy prices are the highest. The pumping to replenish the upper reservoir is performed during off-peak hours when electricity costs are low. Then they are released through the power plant when the electricity prices are higher. The system makes money even though it consumes more energy. 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 24 h (DOE 2002).

Run-of-river projects use the natural flow of the river and produce relatively little change in the stream channel and stream flow. A peaking project impounds and releases 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 flow rate through the turbines and the hydraulic head. The hydraulic head is the elevation difference the water falls (drops) in passing through the plant or to the tailwater, which ever elevation difference is less. Project design may concentrate on either of these flow and head variables or both, and on the hydropower plant installed designed 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 duration of the period; the plant factor (the fraction of time 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 meter of water, weighing 10^3 kg, falling a distance of 1 m, acquires 9.81×10^3 J (nm) 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 meters per second falling a height of H_t meters 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 given a head of H_t and 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

$$KWH_t = 9.81 \, q_t H_t (\text{seconds in period } t)/$$
$$(\text{seconds per hour})$$
$$= 9.81 \, q_t H_t (\text{seconds in period } t)/3600$$
$$(11.56)$$

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 (fraction) of e equals

$$KWH_t = 9.81 \, Q_t^T \, H_t \, e/3600$$
$$= 0.002725 \, Q_t^T \, H_t \, e \qquad (11.57)$$

The energy required for pumped storage, where instead of producing energy the turbines are used to pump water up to a higher level, is

$$KWH_t = 0.002725 \, Q_t^T \, H_t/e \qquad (11.58)$$

For Eqs. 11.57 and 11.58, Q_t^T is expressed in cubic meters and H_t is in meters. 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 (and possibly the reservoir release if the tailwater elevation affects the head). 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 model 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 number of model variables and constraints. For a preliminary screening of hydropower capacities prior to a more detailed analysis (e.g., using simulation or other nonlinear 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, and (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 average heads, H_t^o, and flows, Q_t^o, can be used in a linear approximation of the non-separable product terms, $Q_t^T H_t$.

$$Q_t^T H_t = H_t^o Q_t^T + Q_t^o H_t - Q_t^o H_t^o \qquad (11.59)$$

Again, the model may need to be solved several times in order to identify reasonably accurate average flow and head estimates, Q_t^o and H_t^o, in each period t.

The amount of electrical energy produced is limited by the installedkilowatts of plant capacity P as well as on the plant factor p_t. The plant factor is a measure of hydroelectric power plant use in each time period. 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.

$$\mathrm{KWH}_t \leq P \, h_t p_t \qquad (11.60)$$

11.8   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 iirrigation. These uses generally require the withdrawals of water from a river system, from reservoirs, or from other surface or groundwater bodies. The water withdrawn may be only partially consumed, and that which is not consumed may be returned, 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 (1) reservoir storage, possibly for recreational use as well as for water supply; (2) for flow augmentation, possibly for water quality control or for navigation or for ecological benefits; and (3) for 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, that site and in that period.

$$q_t^s \leq Q_t^s \qquad (11.61)$$

The quantity of water that any particular user expects to receive in each particular period is

termed the *target allocation.* Given a multi-period (e.g., annual) known or unknown target allocation T^s at site s, some (usually known) fraction, f_t, of that target allocation will be expected in 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.

$$q_t^s = f_t^s T^s - D_t^s + E_t^s \qquad (11.62)$$

Even though allowed, one would not expect a solution to contain nonzero 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 (Chap. 9). For example, the benefits derived from the allocation of water for hydropower production in period t in some cases will 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.

$$q_t^s \geq f_t^s T^s \quad \text{for all relevant } t \qquad (11.63)$$

If, for any reason, an allocation variable value q_t^s must be low, or even zero, due to other more beneficial uses, then clearly from Eq. 11.63 the annual or growing season target allocation T^s would be low (or 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 Chap. 10.

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.

11.9 Lake-Based Recreation

Recreation benefits derived from natural lakes as well as reservoirs are usually dependent on their storage levels. Where recreational facilities have been built, recreational benefits will also be dependent on recreational target lake levels as well. If docks, boathouses, shelters, and other recreational facilities were installed based on some assumed (target) 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 modeled similar to Eq. 11.62. The actual storage volume at the beginning of a recreation period t equals the target storage volume less any deficit or plus any excess.

$$S_t^s = T^s - D_t^s + E_t^s \qquad (11.64)$$

The recreational benefits in any recreational period t can be defined based on 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 (Chap. 9).

11.10   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 Fig. 11.26.

One of the first tasks in modeling this basin is to identify the actual and potential system components and their interdependencies. This is
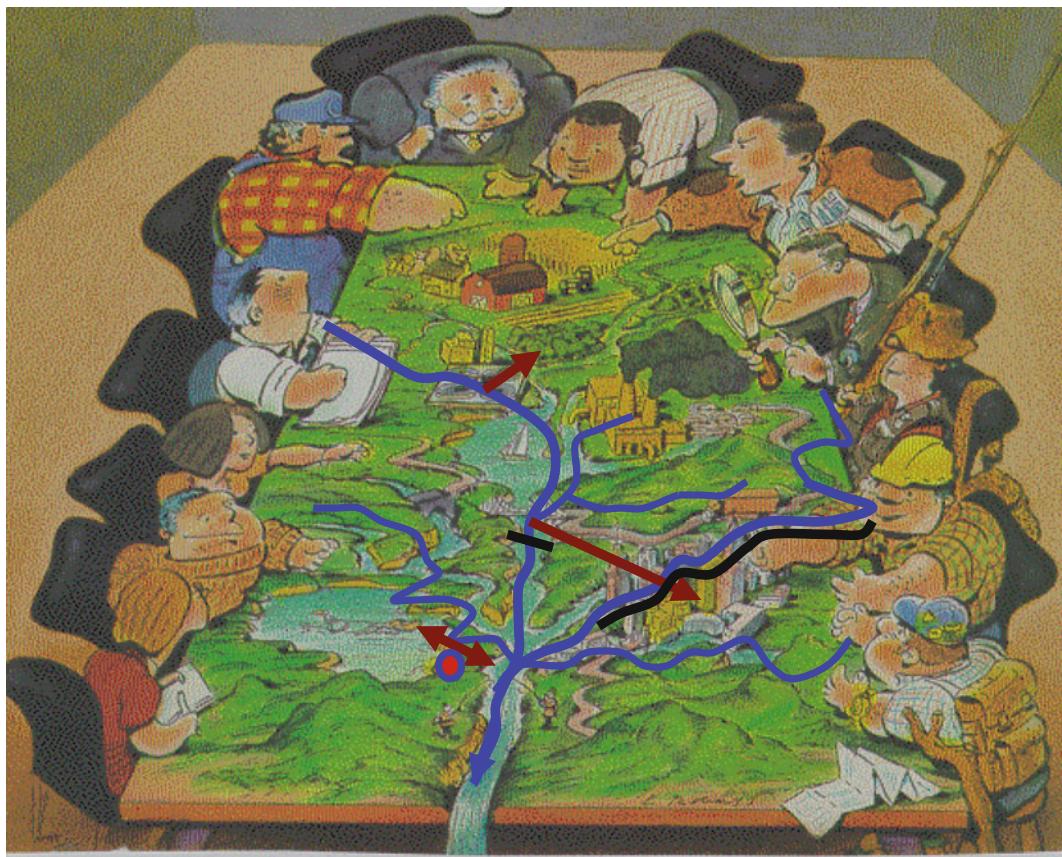
facilitated by drawing a schematic of the system at the level of detail that will address the issues being discussed and of concern to these stakeholders. This schematic can be drawn over the basin as in Fig. 11.27. The schematic without the basin is illustrated in Fig. 11.28.

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.

As shown in the schematic in Fig. 11.28, this river has one streamflow gage 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
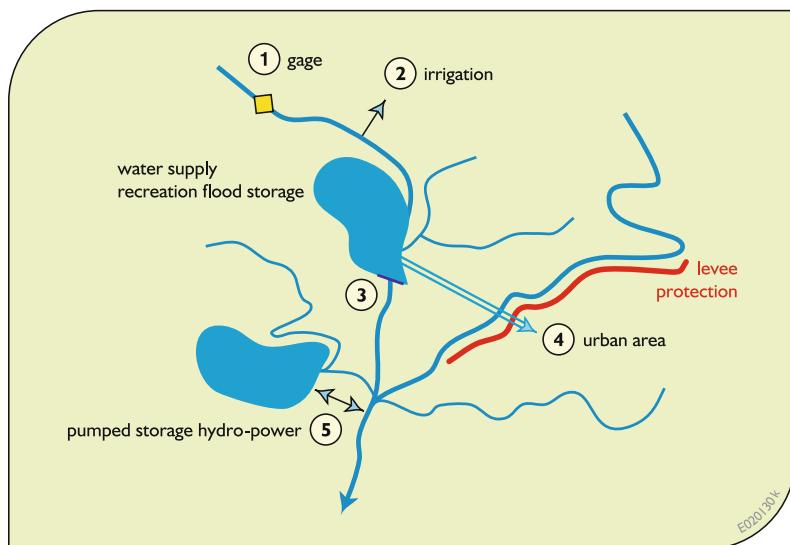
Fig. 11.26   A multipurpose river system whose management is of concern to numerous stakeholders

Fig. 11.27 A schematic representation of the basin components and their interdependencies drawn over the map image of the basin

Fig. 11.28 Schematic of river system showing components of interest at designated sites

flood control. The downstream reservoir is strictly for hydropower production.

Before developing a model of this river system, the number of 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 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 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 the periods and the duration of each period will depend on the variation in the hydrology, the demands, and on the particular objectives, as previously discussed.

Once the number and duration of the time steps to be modeled 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 11.4.

For this example assume we are interested in maximizing a weighted combination of all the net economic benefits derived from all the designated uses of water. There could, and no doubt should, be other objective components defined as well, as discussed in Chap. 9. Nevertheless these economic objective components serve the purpose here of illustrating how a model of this river system can be constructed:

The overall objective might be a weighted combination of all net benefits, NB^s, obtained at each site s in the basin:

$$\text{Maximize} \sum_s w_s NB^s \qquad (11.65)$$

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 by how much, may matter. If it is known how much of each of the net benefits derived from each site are to be allocated to each stakeholder group i, then these allocated fraction, f_i, of the total net benefits, NB^s, can be included in the overall objective:

$$\text{Maximize} \sum_i w_i \sum_s f_i NB^s \qquad (11.66)$$

Using methods discussed in Chap. 9, solving the model for various assumed values of these weights can help identify the tradeoffs between different conflicting objectives, Eq. 11.65, or conflicting stakeholder interests, Eq. 11.66.

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 the site index is shown as a superscript.

At site 1:

- No constraints are needed. It is the gage site.

At site 2:

- the diverted water, $X^2(t)$, cannot exceed the streamflow, $Q^2(t)$, at that site.

$$Q^2(t) \ge X^2(t) \quad \forall t \text{ in the irrigation season} \tag{11.67}$$

- the diversion flow, $X^2(t)$, cannot exceed the diversion channel capacity, X^2.

$$X^2 \ge X^2(t) \quad \forall t \text{ in the irrigation season} \tag{11.68}$$

Table 11.4 Names associated with required variables and functions at each site in Fig. 11.28

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, δ_t^2
irrigation diversion channel capacity, X^2

site 3

active initial storage volume. $S^3(t)$, in each period t
dead storage volume, Kd^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, δ_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

The units of these variables and parameters, however defined, must be consistent

- the diversion flow, $X^2(t)$, must meet the irrigation target, $\delta_t^2 T^2$

$$X^2(t) \geq \delta_t^2 T^2 \quad \forall t \text{ in the irrigation season}$$
$$(11.69)$$

- NB^2 = benefit function associated with the annual target irrigation allocation, T^2, less the annual cost function associated with the diversion channel capacity, X^2.

At site 3:

- storage volume mass balances (continuity of storage), assuming no losses.

$$S^3(t+1) = S^3(t) + Q^3(t) - X^3(t) - R^3(t)$$
$$\forall t, \ T + 1 = 1$$
$$(11.70)$$

- define storage deficits, $D^3(t)$, and excesses, $E^3(t)$, relative to recreation target, T^3.

$$S^3(t) = T^3 - D^3(t) + E^3(t)$$
$$\forall t \text{ in recreation season plus} \quad (11.71)$$
$$following \text{ period.}$$

- diverted water, $X^3(t)$, cannot exceed diversion channel capacity, X^3.

$$X^3(t) \leq X^3 \quad \forall t \qquad (11.72)$$

- reservoir storage capacity constraints involving dead storage, K_D^3, and flood storage, K_F^3, capacities.

$$S^3(t) \leq K^3 - K_D^3 - K_F^3$$
$$\forall t \text{ in flood season plus following period.}$$
$$S^3(t) \leq K^3 - K_D^3$$
$$\text{for all other periods } t.$$
$$(11.73)$$

- NB^3 = sum of annual benefit functions for T^3 and K_F^3 less annual costs of K^3 and X^3 less annual recreation losses associated with all $D^3(t)$ and $E^3(t)$.

At site 4:

- define deficit diversion, $D^4(t)$, from site 3, associated with target, $\delta_t^4 T^4$, if any.

$$X^3(t) = \delta_t^4 T^4 - D^4(t) \quad \forall t \qquad (11.74)$$

- channel capacity, Q^4, must equal peak flood flow, PQ_t^4, associated with selected return period, T.

$$Q^4 = PQ_T^4 \qquad (11.75)$$

- NB^4 = sum of annual benefit functions for T^4 less annual cost of Q^4 less annual losses associated with all $D^4(t)$.

At site 5:

- continuity of pumped storage volumes, involving inflows, $QI^5(t)$, and outflows, $QO^5(t)$, and assuming no losses.

$$S^5(t+1) = S^5(t) + QI^5(t) - QO^5(t) \quad \forall t$$
$$(11.76)$$

- active storage capacity involving dead storage, K_D^5.

$$S^5(t) \leq K^5 - K_D^5 \quad \forall t \qquad (11.77)$$

- pumped inflows, $QI^5(t)$, 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, $Q^5(t) - Q^3(t)$.

$$QI^5(t) \leq Q^5(t) - Q^3(t) + R^3(t) \quad \forall t$$
$$(11.78)$$

- define the energy produced, $EP^5(t)$, given the average storage head, $H(t)$, flow through the turbines, $QO^5(t)$, and efficiency, e.

$$EP^5(t) = (\text{const.})(H(t))(QO^5(t))e \quad \forall t$$
$$(11.79)$$

- define the energy consumed, $EC^5(t)$, from pumping given the amount pumped, $QI^5(t)$.

$$EC^5(t) = (\text{const.})(H(t))(QI^5(t))/e \quad \forall t$$
$$(11.80)$$

- Energy production, $EP^5(t)$, and consumption, $EC^5(t)$, constraints given power plant capacity, P^5.

$$EP^5(t) \le P^5 \text{ (hours of energy production in } t)$$
$$\forall t$$
$$(11.81)$$

$$EC^5(t) \le P^5 \text{(hours of pumping in } t) \quad \forall t$$
$$(11.82)$$

- Define the average storage head, $H(t)$, based on storage head functions, $h(S^5(t))$.

$$H(t) = \left[h(S^5(t+1)) + h(S^5(t))\right]/2 \quad \forall t$$
$$(11.83)$$

- NB^5 = Sum of benefits for the energy produced, $EP^5(t)$, less the costs of the energy consumed, EC^5(t), less the annual costs of capacities K^5 and P^5.

Equations 11.67–11.83 together with objective Eq. 11.65 or 11.66 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 Chaps. 4 and 9.

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 surely will 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 nonlinear or dynamic programming (optimization) methods are more difficult to use.

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 modeling framework. This can be done as outlined in Sect. 5.4 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 the range of system design and operation policies that should be further analyzed using more detailed analyses. Optimization models of the type just developed serve as ways to eliminate inferior alternatives from further consideration more than as ways of finding a solution all stakeholders will accept as the best.

11.11   Project Scheduling

The river basin models discussed thus far in this chapter deal with static planning situations in that system components and their capacities once determined are not assumed to change over time. Project capacities, targets, and operating policies take on fixed values and one examines "snapshot steady-state" solutions for a particular time in the future. These "snapshots" only allow for fluctuations caused by the variability of supplies and demands. The non-hydrologic world is seldom static, however. Targets and goals and policies 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 10-, 20-, or 30-year period. But static models may identify projects for implementation in early years which in later year simulations do not appear in the solutions (Chap. 4 contains an example of this).

This is the common problem in capacity expansion, where each project has a fixed construction and implementation cost as well as variable operating, repair, and maintenance cost component. 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 hydrologic 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 projectscheduling 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, i.e., what other projects are present and how they are designed and operated.
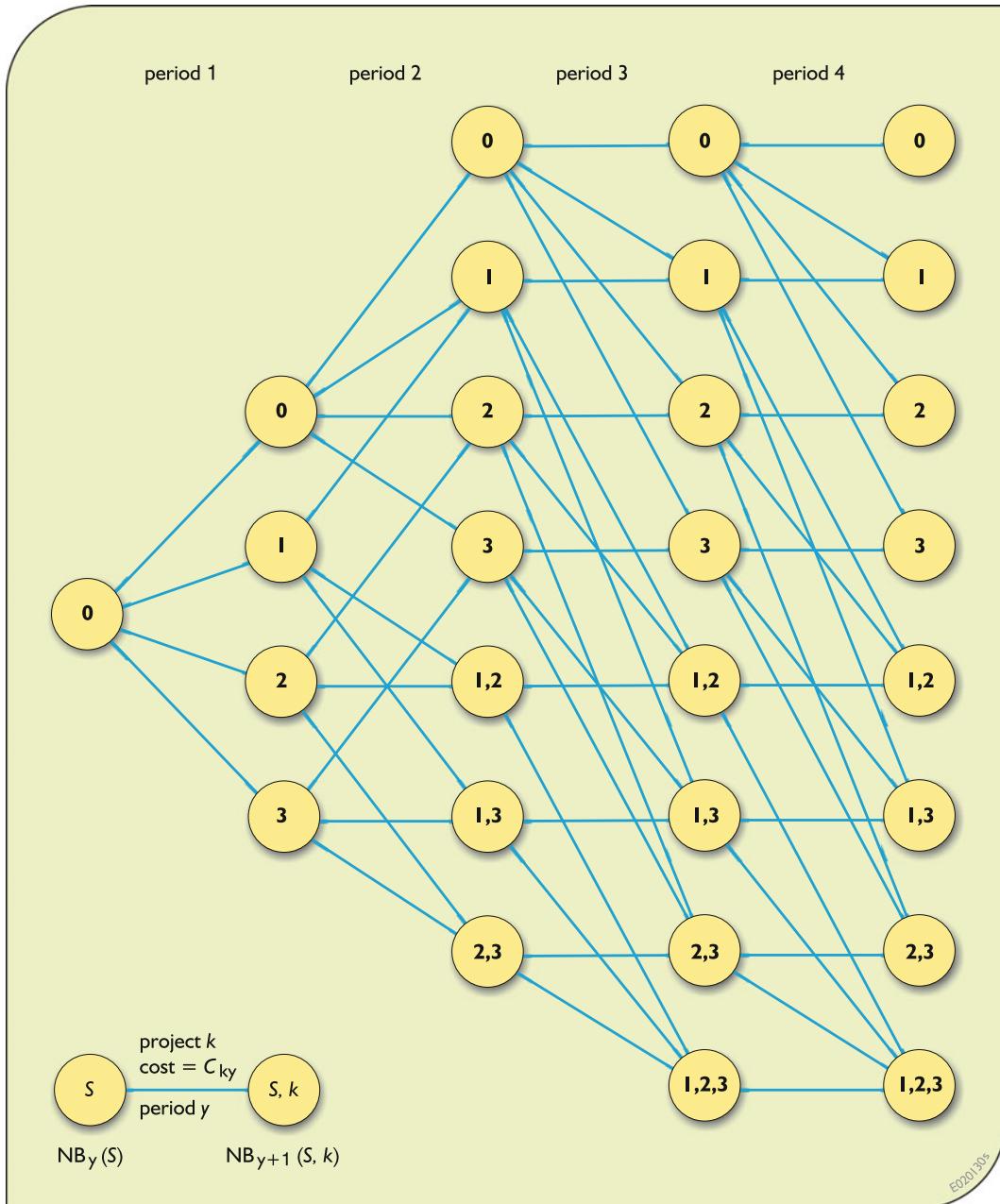
Consider a situation in which n fixed-scale discrete projects may be built during the planning period. The scheduling problem is to determine which of the projects to build and in what order. The solution of this problem generally requires a resolution of the timing problem. When should each project be built, if at all?

For example, assume there are $n = 3$ discrete projects that might be beneficial to implement sometime over the next 20 years. Let this 20-year period consists of four 5-year construction periods y. The actual benefits derived from any new project may depend on the projects that already exist. Let S be the set of projects existing at the beginning of any construction period. Finally let $Na_y(S)$ be the maximum present value of the total net benefits derived in construction period y associated with the projects in the set S. Here "benefits" refer to any composite of system performance measures.

These benefit values for various combinations 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 5-year periods. It might be possible to just do this for 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 Fig. 11.29, the nodes of this network represent the sets S of projects that exist at the beginning of the construction period. For these sets S we have the present value of their benefits, $NB_y(S)$, in the next 5-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

Fig. 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

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 optimal is to find the best (maximum benefits less costs) paths through the network. Each link represents a net benefit, $NB_y(S)$ over the next 5-years obtained from the set of

projects, S, that exist less the cost of adding a new project k.

Using linear programming, one can define a continuous nonnegative 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, 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 (that equals the appropriate $NB_y(S)$ value), and 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.

$$\text{Maximize} \sum_i \sum_j (NB_i - C_{ij})X_{ij} \quad (11.84)$$

Subject to
Continuity at each node

$$\sum_h X_{hi} = \sum_j X_{ij} \quad (11.85)$$
for each node i in the network.

Sum of all decision variable values on the links in any one period y must be 1. For example in period 1

$$X_{00} + X_{01} + X_{02} + X_{03} = 1 \quad (11.86)$$

The sums in Eq. 11.86 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 Chap. 4). For a backward moving solution procedure, let

s = subset of projects k not contained in the set S ($s \notin S$).

$\$_y$ = the maximum project implementation funds available in period y.

$F_y(S)$ = the present value of the total benefits over the remaining periods, y, $y + 1$, …, 4.

$F_{Y+1}(S) = 0$ 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

$$F_y(S) = \text{maximum} \left\{ NB_y(S) - \sum_{k \in s} C_{ky} + F_{y+1}(S+s) \right\}$$
$$\text{for all } S \quad s \notin S \quad \sum_{k \in s} C_{ky} \leq \$_y$$
$$(11.87)$$

Defining $F_y'(S)$ as the present value of the total benefits of all new projects in the set S implemented in all periods up to and including period y, and the subset s of projects k being considered in period y now belonging to the set 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

$$F_y'(S) = \text{maximum} \left\{ NB_y(S - s) - \sum_{k \in s} C_{k,y} + F_{y-1}'(S-s) \right\}$$
$$\text{for all } S \quad s \in S \quad \sum_{k \in s} C_{ky} \leq \$_y$$
$$(11.88)$$

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 constriction 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 Fig. 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 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.

This project scheduling problem by its very nature must deal with uncertainty. A relatively recent contribution to this literature is the work of Haasnoot et al. (2013), Walker et al. (2013).

11.12 Conclusions

This chapter on river basin planning models introduces some ways of modeling 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 Chaps. 7 and 8, these uncertainties may have a substantial effect on 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, can be of value prior to committing to more costly design modeling exercises.

Preliminary screening of river basin systems, especially given multiple objectives, is a challenge to accomplish in an efficient and effective manner. The modeling methods and approaches discussed in this chapter serve as an introduction to that art.

References

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global environmental Change, 23*, 485–498.

Maidment, D. R. (1993). *Handbook of hydrology*. NY: McGraw-Hill Inc.

Rippl, W. (1883). The capacity of storage reservoirs for water supply. In *Proceedings of the Institute of Civil Engineers (Brit.)* (Vol. 71).

U.S. DOE. (2002, April). *Hydropower program*. http://hydropower.inel.gov or http://hydropower.id.doe.gov/

USACE. (1991). Hydrology and hydraulics workshop on Riverine Levee Freeboard, Monticello, Minnesota, Report SP-24, Hydrologic Engineering Center, US Army Corps of Engineers, 272 pp.

Walker, W. E., Haasnoot, M., & Kwakkel, J. H. (2013). Adapt or Perish: A review of planning approaches for adaptation under deep uncertainty. *Sustainability, 5*, 955–979.

Additional References (Further Reading)

Becker, L., & Yeh, W. W. G. (1974). Timing and sizing of complex water resource systems. *Journal of the Hydraulics Division, ASCE , 100*(HYIO), 1457–1470.

Becker, L., & Yeh, W. W.-G. (1974a). Optimal timing, sequencing, and sizing of multiple reservoir surface water supply facilities. *Water Resources Research, 10*(1), 57–62.

Bedient, P. B., & Huber, W. C. (1992). *Hydrology and floodplain analysis* (2nd ed.). Reading: Addison Westley.

Basson, M. S., Allen, R. B., Pegram, G. G. S., & van Rooyen, J. A. (1994). *Probabilistic management of water resource and hydropower systems*. Highlands Ranch: Water Resources Publications.

Chin, D. A. (2000). *Water-resources engineering*. Upper Saddle River: Prentice Hall.

Erlenkotter, D. (1973a). Sequencing of interdependent hydroelectric projects. *Water Resources Research, 9*(1), 21–27.

Erlenkotter, D. (1973b). Sequencing expansion projects. *Operations Research, 21*(2), 542–553.

Erlenkotter, D. (1976). Coordinating scale and sequencing decisions for water resources projects. *Economic Modeling for Water Policy Evaluation, NorthHolland/TIMS Studies in the Management Sciences, 3*, 97–112.

Erlenkotter, D,. & Rogers, J. S. (1977). Sequencing competitive expansion projects. *Operations Research*, *25*(6), 937–951.

Hall, W. A., Tauxe, G. W., & Yeh, W. W.-G. (1969). An alternative procedure for optimization of operations for planning with multiple river, multiple purpose systems. *Water Resources Research, 5*(6), 1367–1372.

Jacoby, H. D., & Loucks, D. P. (1972). Combined use of optimization and simulation models in river basin planning. *Water Resources Research, 8*(6), 1401–1414.

James, L. D., & Lee, R. R. (1971). *Economics of water resources planning*. New York: McGraw-Hill Book Company.

Linsley, R. K., Franzini, J. B., Freyberg, D. L., & Tchobanoglous, G. (1992). *Water-resources engineering*. NY: McGraw-Hill Inc.

Loucks, D. P. (1976). Surface water quantity management, Chap. 5. In A. K. Biswas (Ed.), *Systems approach to water management*. New York: McGraw-Hill Book Company.

Loucks, D. P., Stedinger, J. R., & Haith, D. A. (1981). *Water resource systems planning and analysis*. Englewood Cliffs: Prentice Hall Inc.

Major, D. C., & Lenton, R. L. (1979). *Applied water resources systems planning*. Englewood Cliffs: Prentice Hall Inc.

Maass, A., Hufschmidt, M. M., Dorfman, R., Thomas, H. A., Jr., Marglin, S. A., & Fair, G. M. (1962). *Design of water resource systems*. Cambridge: Harvard University Press.

Mays, L. W. (2001). *Water resources engineering*. NY: Wiley.

Mays, L. W., & Tung, Y.-K. (1992). *Hydrosystems engineering and management*. NY: McGraw-Hill Inc.

Morin, T. L. (1973). Optimal Sequencing of Capacity Expansion Projects. *Journal of the Hydraulics Division, ASCE, 99*(HY9), 1605–1622.

Morin, T. L., & 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, 10*(1), 49–50.

National Research Council. (1999). *New strategies for America's Watersheds*. Washington, DC: National Academy Press.

National Research Council. (2000). *Risk analysis and uncertainty in flood damage reduction studies*. Washington, DC: National Academy Press.

O'Laoghaire, D. T., & Himmelblau, D. M. (1974). *Optimal expansion of a water resources system*. New York: Academic Press Inc.

ReVelle, C. (1999). *Optimizing reservoir resources*. NY: Wiley.

Thomas, H. A., Jr., & Burden, R. P. (1963). *Operations research in water quality management*. Cambridge: Harvard Water Resources Group.

USACE. (1991a). Benefit determination involving existing levees. Policy Guidance Letter No. 26, Headquarters, U.S. Army Corps of Engineers, Washington, DC.

USACE. (1999). Risk analysis in geotechnical engineering for support of planning studies, ETL 1110-2-556, Headquarters, U.S. Army Corps of Engineers, Washington, DC.

Viessman, W., Jr., & Welty, C. (1985). *Water management technology and institutions*. NY: Harper & Row, Pubs.

Wurbs, R. A., & James, W. P. (2002). *Water resources engineering*. Upper Saddle River: Prentice Hall.

Exercises

11.1 Using the following information pertaining to the drainage area and discharge in the Han River in South Korea, develop an equation for predicting the natural unregulated flow at any site in the river, by plotting average flow as a function of catchment area. What does the slope of the function equal?

Gage point	Catchment area (km^2)	Average flow (10^6 m^3/year)
First bridge of the Han River	25,047	17,860
Pal Dang dam	23,713	16,916
So Yang dam	2703	1856
Chung Ju dam	6648	4428
Yo Ju dam	10,319	7300
Hong Chun dam	1473	1094
Dal Chun dam	1348	1058
Kan Yun dam	1180	926
Im Jae dam	461	316

11.2 In watersheds characterized by significant elevation changes, one can often develop reasonable predictive equations for average annual runoff per hectare as a function of elevation. Describe how one would use such a function to estimate the natural average annual flow at any gage in a watershed which is marked by large elevation changes and little loss of water from stream channels due to evaporation or seepage.

11.3 Compute the storage-yield function for a single reservoir system by the mass diagram and modified sequent peak methods given the following sequences of annual flows: (7, 3, 5, 1, 2, 5, 6, 3, 4). Next assume that each year has two distinct

hydrologic seasons, one wet and the other dry, and that 80% of the annual inflow occurs in season $t = 1$ and 80% of the yield is desired in season $t = 2$. Using the modified sequent peak method, show the increase in storage capacity required for the same annual yield resulting from within-year redistribution requirements.
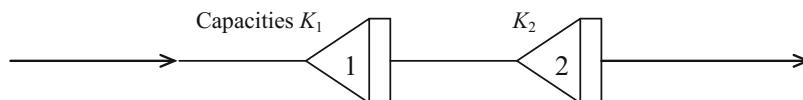
11.4 Write two different linear programming models for estimating the maximum constant reservoir release or yield Y given a fixed reservoir capacity K, and for estimating the minimum reservoir capacity K required for a fixed yield Y. Assume that there are T time periods of historical flows available. How could these models be used to define a storage capacity-yield function indicating the yield Y available from a given capacity K?

11.5(a) Construct an optimization model for estimating the least-cost combination of active storage capacities, K_1 and K_2, of two reservoirs located on a single stream, used to produce a reliable constant annual flow or yield (or greater) downstream of the two reservoirs. Assume that the cost functions $C_s(K_s)$ at each reservoir site s are known and there is no dead storage and no evaporation. (Do not linearize the cost functions; leave them in their functional form.) Assume that 10 years of monthly unregulated flows are available at each site s.

(b) Describe the two-reservoir operating policy that you would incorporate into a model model to check the solution obtained from the optimization model.

Reservoir capacity	Flood stage for flood of return period T					Annual capacity cost
	$T = 1$	$T = 2$	$T = 5$	$T = 10$	$T = 100$	
0	30	105	150	165	180	10[a]
5	30	80	110	120	130	25
10	30	55	70	75	80	30
15	30	40	45	48	50	40
20	30	35	38	39	40	70

[a]10 is fixed cost if capacity > 0; otherwise, it is 0
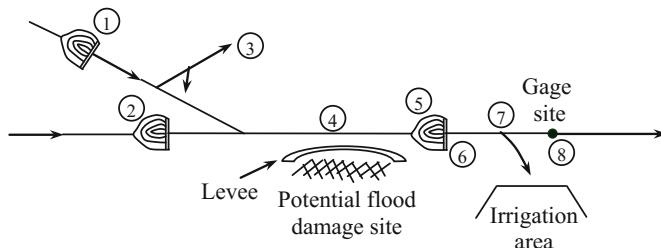
Flood stage	Cost of flood damage
30	0
50	10
70	20
90	30
110	40
130	50
150	90
180	150

11.7 Develop a deterministic, static, within-year model for evaluating the development alternatives in the river basin shown in the accompanying figures. Assume that there are $t = 1, 2, 3, \ldots, n$ within-year periods and that the objective is to maximize the total annual net benefits in the basin. The solution of the model should define the reservoir capacities (active + flood storage capacity), the annual allocation targets, the levee capacity required to protect site 4 from a T-year flood, and the within-year

Capacities K_1 K_2

11.6 Given the information in the accompanying tables, compute the reservoir capacity that maximizes the net expected flood damage reduction benefits less the annual cost of reservoir capacity.

period allocations of water to the uses at sites 3 and 7. Clearly define all variables and functions used, and indicate how the model would be solved to obtain the maximum net benefit solution.

Gage site

Levee Potential flood
damage site

Irrigation area

Site	Fraction of gage flow	
1	0	Potential reservoir for water supply
2	0.3	Potential reservoir for water supply, flood control
3	0.15	Diversion to a use, 60% of allocation returned to river
4	–	Existing development, possible flood protection from levee
5	0.6	Potential reservoir for water supply, recreation
6	–	Hydropower; plant factor = 0.30
7	0.9	Potential diversion to an irrigation district
8	1	Gage site

For simplicity, assume no evaporation losses or dead storage requirements. Omitting the appropriate subscripts t for time periods and s for site, let T, K, D, E, and P be the target, reservoir capacity, deficit, excess, and power plant capacity variables, respectively. Let Q_t and R_t be the natural streamflows and reservoir releases, and S_t be the initial reservoir storage volumes in period t. K_f will denote the flood storage capacity at site 2 that will contain a peak flow of QS and QR is the downstream channel flood flow capacity. The relationship between QS and K_f is defined by the function $\kappa(QS)$. The unregulated design flood peak flow for which protection is required is QN. KWH will be the kilowatt-hours of energy, H will be net storage head, h_t the hours in a period t. The variable q will be the water supply allocation. Benefit functions will be $B()$, $L()$ will denote loss functions and $C()$ will denote the cost functions.

11.8 List the potential difficulties involved when attempting to structure models for defining

(a) Water allocation policies for irrigation during the growing season.
(b) Energy production and capacity of hydroelectric plants.
(c) Dead storage volume requirements in reservoirs.
(d) Active storage volume requirements in reservoir.
(e) Flood storage capacities in reservoirs.
(f) Channel improvements for damage reduction.
(g) Evaporation and seepage losses from reservoirs.
(h) Water flow or storage targets using long-run benefits and short- run loss functions.

11.9 Assume that demand for water supply capacity is expected to grow as $t(60 - t)$, for t in years. Determine the minimum present value of construction cost of some subset of water supply options described below so as to always have sufficient capacity to meet demand over the next 30 years. Assume that the water supply network currently has no excess capacity so that some project must be built immediately. In this problem, assume that project capacities are independent and thus can be summed. Use a discount factor

equal to exp(−0.07 t). Before you start, what is your best guess at the optimal solution?

Project number	Construction cost	Capacity
1	100	200
2	115	250
3	190	450
4	270	700

11.10 (a) Construct a flow diagram for a simulation model designed to define a storage-yield function for a single reservoir given known inflows in each month t for n years. Indicate how you would obtain a steady-state solution not influenced by an arbitrary initial storage volume in the reservoir at the beginning of the first period. Assume that evaporation rates (mm per month) and the storage volume/surface area functions are known.

(b) Write a flow diagram for a simulation model to be used to estimate the probability that any specific reservoir capacity, K, will satisfy a series of known release demands, r_t, downstream given unknown future inflows, i_t. You need not discuss how to generate possible future sequences of streamflows, only how to use them to solve this problem.

11.11 (a) Develop an optimization model for finding the cost-effective combination of flood storage capacity at an upstream reservoir and channel improvements at a downstream potential damage site that will protect the downstream site from a prespecified design flood of return period T. Define all variables and functions used in the model

(b) How could this model be modified to consider a number of design floods T and the benefits from protecting the potential damage site from those design floods? Let BF_T be the annual expected flood protection benefits at the damage site for a flood having return periods of T.

(c) How could this model be further modified to include water supply requirements of A_t to be withdrawn from the reservoir in each month t? Assume known natural flows Q_t^s at each site s in the basin in each month t.

(d) How could the model be enlarged to include recreation benefits or losses at the reservoir site? Let T^s be the unknown storage volume target and D_t^s be the difference between the storage volume S_t^s and the target T^s if $S_t^s - T^s > 0$, and E_t^s be the difference if $T^s - S_t^s > 0$. Assume that the annual recreation benefits $B(T^s)$ are a function of the target storage volume T^s and the losses $L^D(D_t^s)$ and $L^E(E_t^s)$ are associated with the deficit D_t^s and excess E_t^s storage volumes.

11.12 Given the hydrologic and economic data listed below, develop and solve a linear programming model for estimating the reservoir capacity K, the flood storage capacity K_f, and the recreation storage volume target T that maximize the annual expected flood control benefits, $B_f(K_f)$, plus the annual recreation benefits, $B(T)$, less all losses $L^D(D_t)$ and $L^E(E_t)$ associated with deficits D_t or excesses E_t in the periods of the recreation season, minus the annual cost $C(K)$ of storage reservoir capacity K. Assume that the reservoir must also provide a constant release or yield of $Y = 30$ in each period t. The flood season begins at the beginning of period 3 and lasts through period 6. The recreation season begins at the beginning of period 4 and lasts though period 7.

Period	1	2	3	4	5	6	7	8	9
Inflows to reservoir	50	30	20	80	60	20	40	10	70

$$B_f(K_f) = \begin{cases} 12K_f & \text{if} \quad K_f \le 5 \\ 60 + 8(K_f - 5) & \text{if} \quad 5 \le K_f \le 15 \\ 140 + 4(K_f - 15) & \text{if} \quad K_f \le 15 \end{cases}$$

$$C(K_f) = \begin{cases} 45 + 10K & \text{if} \quad 0 \le K_f \le 5 \\ 95 + 6(K_f - 5) & \text{if} \quad 5 \le K_f \le 20 \\ 185 + 10(K_f - 20) & \text{if} \quad 20 \le K_f \le 40 \\ 385 + 15(K - 40) & \text{if} \quad K_f \ge 40 \end{cases}$$

$B(T) = 9T,$ where T is a particular unknown value of -reservoir storage

$L^D(D_t) = 4D_t,$ where D_t is $T - S_t$ if $T \ge S_t$
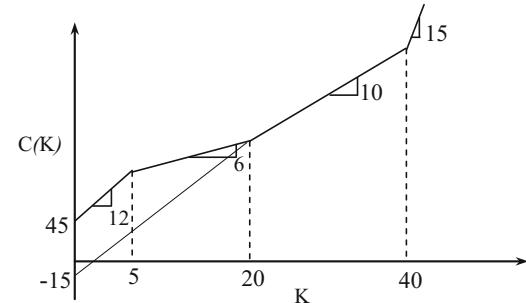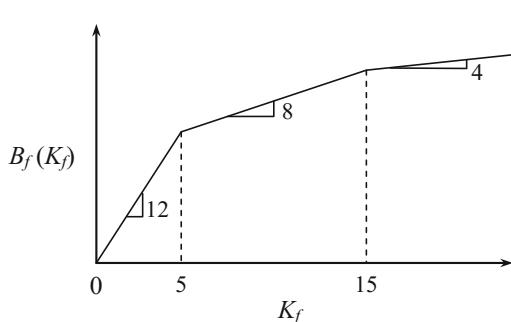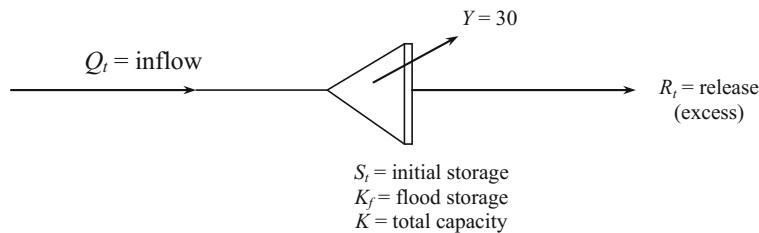
$L^E(E_t) = 2 E_t,$ where E_t is $S_t - T$ if $S_t \ge T$

11.13 The optimal operation of multiple reservoir systems for hydropower production presents a very nonlinear and often difficult problem.

Use dynamic programming to determine the operating policy that maximizes the total annual hydropower production of a two-reservoir system, one downstream of the other. The releases R_{1t} from the upstream reservoir plus the unregulated incremental flow $(Q_{2t} - Q_{1t})$ constitute the inflow to the downstream dam. The flows Q_{1t} into the upstream dam in each of the four seasons along with the incremental flows $(Q_{2t} - Q_{1t})$ and constraints on reservoir releases are given in the accompanying two tables:

Upstream dam (flow in 10^6 m³/period)			
Season t	Inflow Q_{1t}	Minimum release	Maximum release through turbines
1	60	20	90
2	40	30	90

(continued)

$Y = 30$

$Q_t = $ inflow

$R_t = $ release (excess)

$S_t = $ initial storage
$K_f = $ flood storage
$K = $ total capacity

Upstream dam (flow in 10^6 m³/period)			
Season t	Inflow Q_{1t}	Minimum release	Maximum release through turbines
3	80	20	90
4	120	20	90

Downstream dam (flow in 10^6 m³/period)			
Season t	Incremental flow, $(Q_{2t} - Q_{1t})$	Minimum release	Maximum release through turbines
1	50	30	140
2	30	40	140
3	60	30	140
4	90	30	140

Note that there is a limit on the quantity of water that can be released through the turbines for energy generation in any season due to the limited capacity of the power plant and the desire to produce hydropower during periods of peak demand.

Additional information that affects the operation of the two reservoirs are the limitations on the fluctuations in the pool levels (head) and the storage-head relationships:

Data	Upstream dam	Downstream dam
Maximum head, H^{max} (m)	70	90
Minimum head, H^{min} (m)	30	60
Maximum storage volume, S^{max} (m³)	150×10^6	400×10^6
Storage-net head relationship	$H = H^{max}(S/S^{max})^{0.64}$	$H = H^{max}(S/S^{max})^{0.62}$

In solving the problem, discretize the storage levels in units of 10×10^6 m³. Do a preliminary analysis to determine how large a variation in storage might occur at each reservoir. Assume that the conversion of potential energy equals to the product R_iH_i to electric energy is 70% efficient independent of R_i and H_i. In calculating the energy produced in any season t at reservoir i, use the average head during the season

$$\bar{H}_i = \frac{1}{2}[H_i(t) + H_i(t+1)]$$
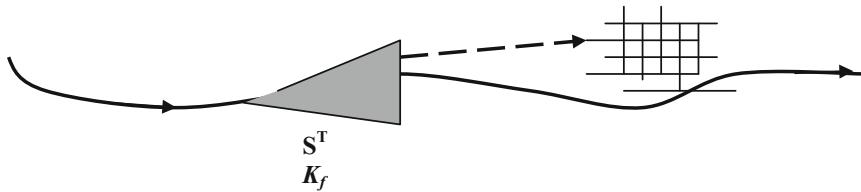
Report your operating policy and the amount of energy generated per year. Find another feasible policy and show that it generates less energy than the optimal policy.

Show how you could use linear programming to solve for the optimal operating policy by approximating the product term $R_i\bar{H}_i$ by a linear expression.

11.14 You are responsible for planning a project that may involve the building of a reservoir to provide water supply benefits to a municipality, recreation benefits associated with the water level in the lake behind the dam, and flood damage reduction benefits. First you need to determine some design variable values, and after doing that you need to determine the reservoir operating policy.

The design variables you need to determine include

- the total reservoir storage capacity (K),
- the flood storage capacity (K_f) in the first season that is the flood season,
- the particular storage level where recreation facilities will be built, called the storage target (S^T) that will apply in seasons 3, 4, and 5—the recreation seasons and finally,
- the dependable water supply or yield (Y) for the municipality.

$$S^T$$
$$K_f$$

Assume you can determine these design variable values based on average flows at the reservoir site in six seasons of a year. These average flows are 35, 42, 15, 3, 15, and 22 in the seasons 1–6, respectively.

The objective is to design the system to maximize the total annual net benefits derived from

- flood control in season 1,
- recreation in seasons 3–5, and
- water supply in all seasons,

  less the annual cost of the

- reservoir and
- any losses resulting from not meeting the recreation storage targets in the recreation seasons.

The flood benefits are estimated to be $2\,K_f^{\,0.7}$.

The recreation benefits for the entire recreation season are estimated to be $8\,ST$.

The water supply benefits for the entire year are estimated to be $20\,Y$.

The annual reservoir cost is estimated to be $3\,K^{1.2}$.
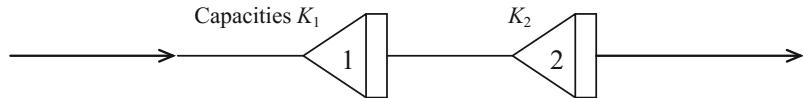
The recreation loss in each recreation season depends on whether the actual storage volume is lower or higher than the storage target. If it is lower the losses are 12 per unit average deficit in the season, and if they are higher the losses are 4 per unit average excess in the season. It is possible that a season could begin with a deficit and end with an excess, or vice versa.

Develop and solve a nonlinear optimization model for finding the values of each of the design variables: K, K_f, ST, and Y and the maximum annual net benefits. (There will be other variables as well. Just define what you need and put it all together in a model.)

Does the solution give you sufficient information that would allow you to simulate the system using a sequence of inflows to the reservoir that are different than the ones used to get the design variable values? If not how would you define a reservoir operating policy? After determining the system's design variable values using optimization, and then determining the reservoir operating policy, you would then simulate this system over many years to get a better idea of how it might perform.

11.15 Suppose you have 19 years of monthly flow data at a site where a reservoir could be located. How could you construct a model to estimate what the required over-year and within year storage needed to produce a specified annual yield Y that is allocated to each month t by some known fraction δ_t. What would be the maximum reliability of those yields? If you wanted to add to that an additional secondary yield having only 80% reliability, how would the model change? Make up 19 annual flows and assume that the average monthly flows are specified fractions of those annual flows. Just using these annual flows and the average monthly fractions, solve your model.

11.16

Capacities K_1 K_2

(a) Develop an optimization model for estimating the least-cost combination of active storage K^1 and K^2 capacities at two reservoir sites on a single stream that are used to produce a reliable flow or yield downstream of the downstream reservoir. Assume 10 years of monthly flow data at each reservoir site. Identify what other data are needed.

(b) Describe the two-reservoir operating policy that could be incorporated into a simulation model to check the solution obtained from the optimization model.

Define

$C^s(K^s)$	cost of active storage capacity at site s; where $s = 1, 2$
K_d^s	dead storage capacity of reservoir at site s; $K_d^s = 0$
S_t^s	storage volume at beginning of period t at site s
L^s	loss of water due to evaporation at site s; $L^s = 0$
R_t^{12}	release from reservoir at site 1 to site 2 in period t
Y_t	yield to downstream in period t
Q_t^s	10 years of monthly natural flows available at each site s
a_o^s	area associated with dead storage volume at site s
a^s	area per unit storage volume at site s
e_t^s	evaporation depth in period t at site s

11.17 Given inflows to an effluent storage lagoon that can be described by a simple first-order Markov chain in each of

T periods t, and an operating policy that defines the lagoon discharge as a function of the initial volume and inflow, indicate how you would estimate the probability distribution of lagoon storage volumes.

11.18 (a) Using the inflow data in the table below, develop and solve a yield model for estimating the storage capacity of a single reservoir required to produce a yield of 1.5 that is 90% reliable in both of the two within-year periods t, and an additional yield of 1.0 that is 70% reliable in period $t = 2$.

(b) Construct a reservoir-operating rule that defines reservoir release zones for these yields.

(c) Using the operating rule, simulate the 18 periods of inflow data to evaluate the adequacy of the reservoir capacity and storage zones for delivering the required yields and their reliabilities. (Note that in this simulation of the historical record the 90% reliable yield should be satisfied in all the 18 periods, and the incremental 70% reliable yield should fail only two times in the 9 years.)

(d) Compare the estimated reservoir capacity with that which is needed using the sequent peak procedure.

Year	Period	Inflow
1	1	1.0
	2	3.0
2	1	0.5
	2	2.5
3	1	1.0
	2	2.0

Year	Period	Inflow
4	1	0.5
	2	1.5
5	1	0.5
	2	0.5
6	1	0.5
	2	2.5
7	1	1.0
	2	5.0
8	1	2.5
	2	5.5
9	1	1.5
	2	4.5

11.19 One possible modification of the yield model of would permit the solution algorithm to determine the appropriate failure years associated with any desired reliability instead of having to choose these years prior to model solution. This modification can provide an estimate of the extent of yield failure in each failure year and include the economic consequences of failures in the objective function. It can also serve as a means of estimating the optimal reliability with respect to economic benefits and losses. Letting F_y be the unknown yield reduction in a possible failure year y, then in place of $\alpha_{py}Y_p$ in the over-year continuity constraint, the term $(Y_p - F_y)$ can be used. What additional constraints are needed to ensure (1) that the average shortage does not exceed $(1 - \alpha_{py})Y_p$ or (2) that at most there are f failure years and none of the shortages exceed $(1 - \alpha_{py})Y_p$.

11.20 In Indonesia there exists a wet season followed by a dry season each year. In one area of Indonesia all farmers within an irrigation district plant and grow rice during the wet season. This crop brings the farmer the largest income per hectare; thus they would all prefer to continue growing rice during the dry season. However, there is insufficient water during the dry season for irrigating all 5000 ha of available irrigable land for rice production. Assume an available irrigation water supply of 32×10^6 m^3 at the beginning of each dry season, and a minimum requirement of 7000 m^3/ha for rice and 1800 m^3/ha for the second crop.

(a) What proportion of the 5000 ha should the irrigation district manager allocate for rice during the dry season each year, provided that all available hectares must be given sufficient water for rice or the second crop?

(b) Suppose that crop production functions are available for the two crops, indicating the increase in yield per hectare per m^3 of additional water, up to 10,000 m^3/ha for the second crop. Develop a model in which the water allocation per hectare, as well as the hectares allocated to each crop, is to be determined, assuming a specified price or return per unit of yield of each crop. Under what conditions would the solution of this model be the same as in part (a)?

11.21 Along the Nile River in Egypt, irrigation farming is practiced for the production of cotton, maize, rice, sorghum, full and short berseem for animal production, wheat, barley, horsebeans, and winter and summer tomatoes. Cattle and buffalo are also produced, and together with the crops that require labor, water. Fertilizer, and land area (feddans). Farm types or management practices are fairly uniform, and hence in any analysis of irrigation

policies in this region this distinction need not be made. Given the accompanying data develop a model for determining the tons of crops and numbers of animals to be grown that will maximize (a) net economic benefits based on Egyptian prices, and (b) net economic benefits based on international prices. Identify all variables used in the model.

Known parameters

C_i — miscellaneous cost of land preparation per feddan

P_i^E — Egyptian price per 1000 tons of crop i

P_i^I — international price per 1000 tons of crop i

v — value of meat and dairy production per animal

g — annual labor cost per worker

f^P — cost of P fertilizer per ton

f^N — cost of N fertilizer per ton

Y_i — yield of crop i, tons/feddan

α — feddans serviced per animal

β — tons straw equivalent per ton of berseem carryover from winter to summer

r^w — berseem requirements per animal in winter

s^{wh} — straw yield from wheat, tons per feddan

s^{ba} — straw yield from barley, tons per feddan

r^s — straw requirements per animal in summer

μ_i^N — N fertilizer required per feddan of crop i

μ_i^P — P fertilizer required per feddan of crop i

l_{im} — labor requirements per feddan in month m, man-days

w_{im} — water requirements per feddan in month m, 1000 m^3

h_{im} — land requirements per month, fraction (1 = full month)

Required Constraints. (Assume known resource limitations for labor, water, and land)

(a) Summer and winter fodder (berseem) requirements for the animals.

(b) Monthly labor limitations.

(c) Monthly water limitations.

(d) Land availability each month.

(e) Minimum number of animals required for cultivation.

(f) Upper bounds on summer and winter tomatoes (assume these are known).

(g) Lower bounds on cotton areas (assume this is known).

Other possible constraints

(a) Crop balances.

(b) Fertilizer balances.

(c) Labor balance.

(d) Land balance.

11.22 In Algeria there are two distinct cropping intensities, depending upon the availability of water. Consider a single crop that can be grown under intensive rotation or extensive rotation on a total of A hectares. Assume that the annual water requirements for the intensive rotation policy are 16,00 m^3 per ha, and for the extensive rotation policy they are 4000 m^3 per ha. The annual net production returns are 4000 and 2000 dinars, respectively. If the total water available is 320,000 m^3, show that as the available land area A increases, the rotation policy that maximizes total net income changes from one that is totally intensive to one that is increasingly extensive.

Would the same conclusion hold if instead of fixed net incomes of 4000 and 2000 dinars per hectares of intensive and extensive rotation, the net income depended on the quantity of crop produced? Assuming that intensive rotation produces twice as much produced by extensive rotation, and that the net income per unit of crop Y is defined by the simple linear function $5 - 0.05Y$, develop and solve a linear programming model to determine the optimal rotation policies if A equals 20, 50, and 80. Need this net income or price function be linear to be included in a linear programming model?

Urban Water Systems

<div style="text-align: right">

12

</div>

"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 urban life profoundly in the 20th 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 top five greatest achievements of engineering in the twentieth century. But providing everyone with clean tap water, especially in urban areas, has yet to be achieved, even in developed nations. The world's population is growing by about 80 million people per year, and is predicted to approach 10 billion by 2050. Over 50% of people on our planet today live in urban areas and that percentage will grow. As populations continually move to cities for improved economic opportunities and a higher standard of living and as cities merge to form megacities, the design and management of water becomes an increasingly important part of integrated urban infrastructure planning and management.
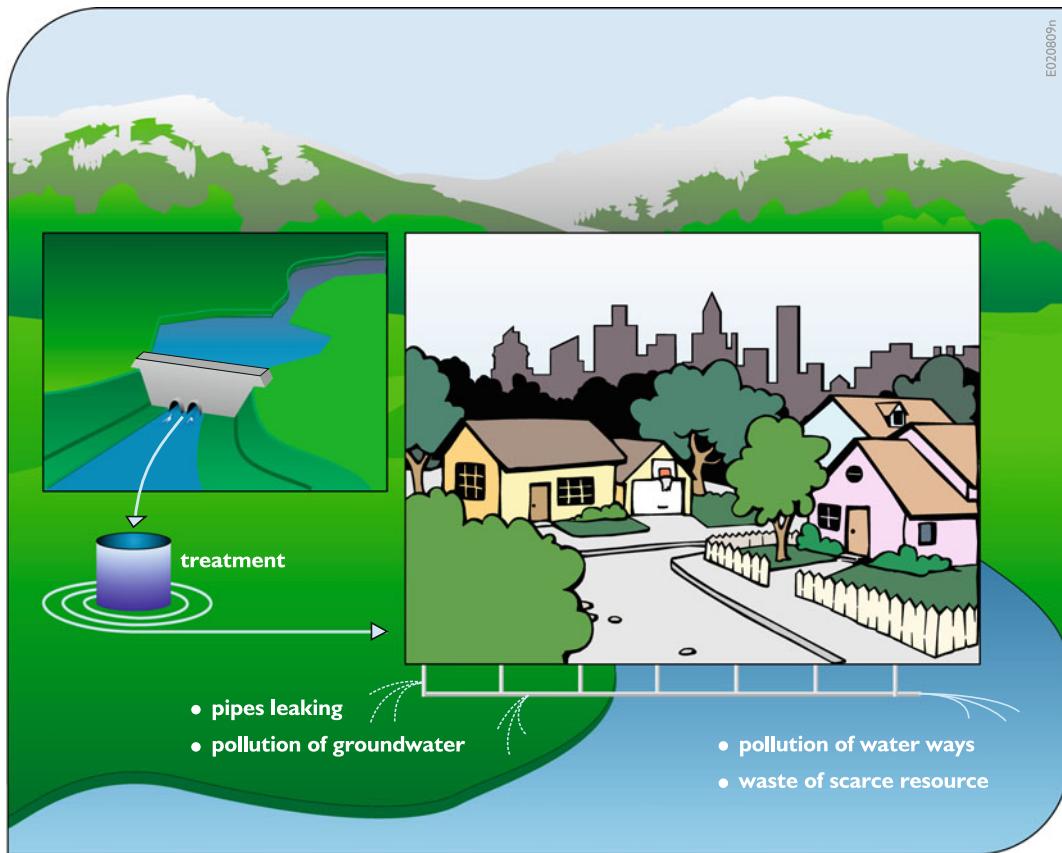
12.1 Introduction

Urban water management involves the planning, design, and operation of infrastructure needed to meet the demands for drinking water and sanitation, the control of infiltration and stormwater runoff, and for recreational parks and the maintenance of urban ecosystems. As urban areas grow, so do the demands for such services. In addition there is an increasing need to make urban water systems more resilient to climate change. All this leads to the realization that urban water management must be an integral part of urban planning in general. Land use decisions impact water supply and wastewater system designs and operation, as well as measures needed for managing stormwater runoff. A functioning urban infrastructure system also requires energy which in turn typically requires water.

Sustainable urban development must focus on the relationships between water, energy, and land use, and often on diversifying sources of water to assure reliable supplies. Integrated urban water management (IUWM) provides both a goal and a framework for planning, designing, and managing urban water systems. It is a flexible process that responds to change and enables stakeholders to participate in, and predict the impacts of, development decisions. It includes the environmental, economic, social, technical, and political aspects of urban water management. It enables better land use planning and the management of its impacts on fresh water supplies, treatment, and distribution; wastewater collection, treatment, reuse and disposal; stormwater collection, use and disposal; and solid waste collection, recycling, and disposal systems. It makes urban development part of integrated basin management oriented toward a more economically, socially, and environmentally sustainable mixed urban–rural landscape.

Fig. 12.1 Schematic showing urban surface water source, water treatment prior to urban use, and some sources of nonpoint urban drainage and runoff and its impacts

While recognizing the need for and the benefits derived from a systems approach to urban planning and development, including its water related components, this chapter will serve as an introduction to each of these components separately, and not all together as a system. This understanding of each component is needed if indeed they eventually will be modeled, designed, and managed as part of the overall urban infrastructure system.

These urban water infrastructure components typically include 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 (sewer) systems and treatment; and urban drainage works. This is illustrated as a simple schematic in Fig. 12.1.

Generic data-driven 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 desire, and many require, high quality potable water. The quality of 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 prior to its use is required. Once treated, water can 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 groundwater leakage and to meet fire protection and other user 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. Wastewater treatment plants remove some of the impurities in the wastewater before discharging it 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 discharge of point and nonpoint pollutants into receiving water bodies can impact the quality of the water in those receiving water bodies. The fate and transport of these pollutants in water bodies can be predicted using water quality models similar to those discussed in Chap. 10.

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 and manage urban water infrastructure systems. The focus of urban water systems modeling 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.

> **Box 12.1 An urban drinking water crisis**
>
> This ongoing (2016) case of urban water management in Flint, Michigan (US) illustrates what can happen even in so-called developed regions if decisions are made without adequate analyses of possible health impacts and the consequences of poor follow-up decisions at various governmental levels.
>
> After a change in April 2014 of the source of the city's drinking water, the city's water distribution system became contaminated with lead. This has created a serious public health danger. As a result, thousands of residents have severely high levels of lead in their blood and are experiencing a variety of serious health problems. Local, state, and federal authorities, and political leaders, did not seem sufficiently concerned until inhabitants of Flint, with the help of others who performed water quality analyses and obtained public health statistics from local hospitals, made it a national issue. In November, 2015, some of the residents filed a federal class action lawsuit against the Governor and 13 other city and state officials. Additional lawsuits have been filed after that and resignations have occurred. In January 2016, the Governor declared the city to be in a state of emergency. Less than 2 weeks later the President of the US declared the drinking water crisis in Flint as a federal emergency authorizing additional help from the federal government.

12.2 Water Treatment

Before water is to be used for human consumption its harmful impurities need to be removed. Communities that do not have adequate water treatment facilities, common in developing regions, often have high incidences of disease and mortality due to contaminated water supplies. A range of syndromes, including acute dehydrating diarrhea (cholera), prolonged febrile illness with abdominal symptoms (typhoid fever), acute bloody diarrhea (dysentery), and chronic diarrhea (Brainerd diarrhea) result in over 3 billion episodes of diarrhea and an estimated 2 million deaths, mostly among children, each year.

Contaminants in natural water supplies can include microorganisms such as Cryptosporidium

and Giardia lamblia, inorganic and organic cancer-causing chemicals (such as compounds containing arsenic, chromium, copper, lead, and mercury), and radioactive materials (such as radium and uranium). As Box 12.1 illustrates, this need to protect water supplies from such contaminants is not limited to developing regions.
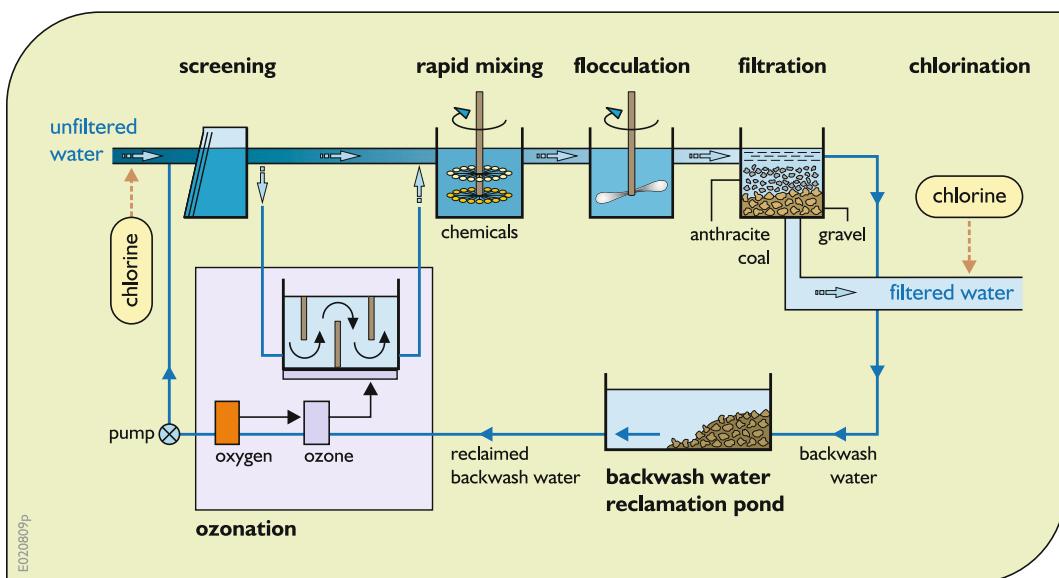
To remove impurities and pathogens, a typical municipal water purification system involves a sequence of physical and chemical processes. Physical and chemical removal processes include initial and final filtering, coagulation, flocculation sedimentation, and disinfection, as illustrated in the schematic of Fig. 12.2.

As shown in Fig. 12.2, one of the first steps in most water treatment plants involves passing raw water through coarse filters to remove sand, grit, and large solid objects. Next, a chemical such as alum is added to the raw water to facilitate coagulation of remaining impurities. As the wastewater 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 trihalomethanes.

Alternatives to chlorine include ozone oxidation (Fig. 12.2). Unlike chlorine, ozone does not stay in the water after it leaves the treatment plant, so it offers 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 it has the same limitation as oxidation. It is ineffective outside the treatment plant.

Fig. 12.2 Typical processes in water treatment plants

Fig. 12.3 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

Figure 12.3 is an aerial view of a water treatment plant serving a population of about 50,000.

12.3   Water Distribution

Water distribution systems include pumping stations, distribution storage, and distribution piping. The hydraulic performance of each component in the distribution network depends upon the performance of other components. Of interest to designers are both the flows and their pressures throughout the network.

The energy at any point within a network of pipes is often represented in three parts: the pressure head, p/γ, 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

center of the pipe. The parameter p is the pressure, e.g., newton per cubic meter (N/m^3), γ 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, and lost from the system such as by 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.

$$\left[p/\gamma + Z + V^2/2g + H_G\right]_{\text{site } i}$$
$$= \left[p/\gamma + Z + V^2/2g + H_L\right]_{\text{site } j} \qquad (12.1)$$
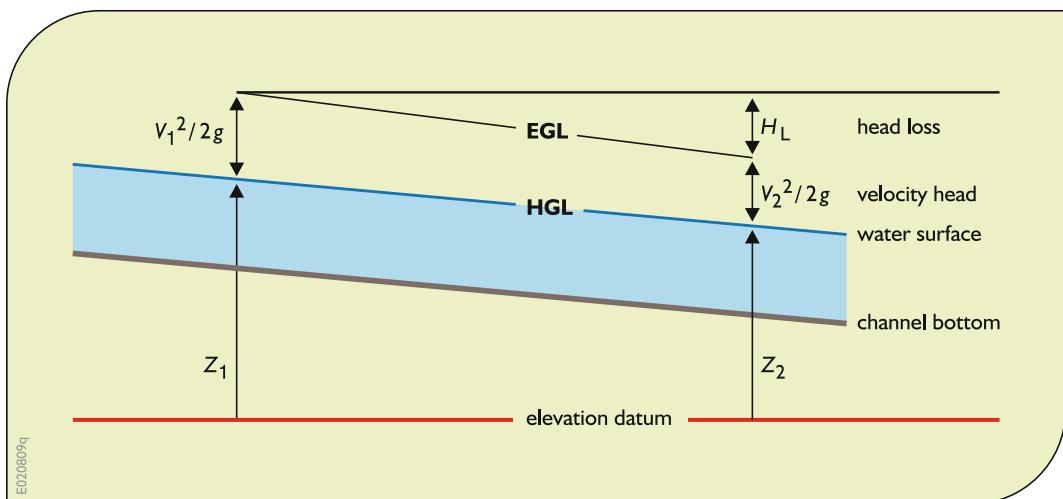
The hydraulic grade is the sum of the pressure head and elevation head $(p/\gamma + Z)$. For open-channel flow the hydraulic grade is the water surface slope, since the pressure head at its surface is 0. For a pressure pipe, the hydraulic head is the height to which a water column would rise in a piezometer—a 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-channel and pressure pipes are illustrated in Figs. 12.4 and 12.5.
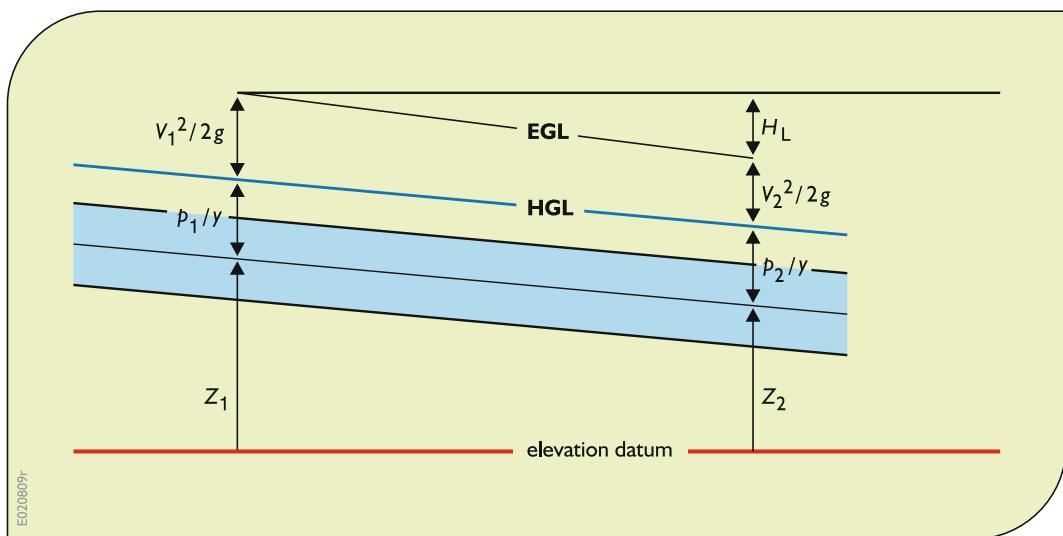
The energy grade is the sum of the hydraulic grade and the velocity head. This is the height to which a column of water would rise in a Pitot tube, but also accounting for fluid velocity. When plotted in profile, as in Fig. 12.5, this is often referred to as the energy grade line, or EGL. At a lake or 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

Fig. 12.4 The energy components along an open channel

Fig. 12.5 The energy components along a pressure pipe

more than one depth having 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 that is commonly used for open-channel flow and Chezy's or Kutter's equation, Hazen–Williams equation, and Darcy–Weisbach or Colebrook–White equations that 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, and the friction or energy slope, $S = H_L/$ Length.

$$V = kCR^x S^y \qquad (12.2)$$

The terms k, x, and y of Eq. 12.2 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., Mays 2000, 2001; Chin 2000).

12.3.1   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.

$$V = \left(R^{2/3}S^{1/2}\right)/n \qquad (12.3)$$

$$Q = AV \qquad (12.4)$$

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 Eq. 12.5. For open-channel flow the pressure heads are 0. Thus for a channel containing water flowing from site i to site j:

$$\left[Z + V^2/2g\right]_{\text{site }i} = \left[Z + V^2/2g + H_L\right]_{\text{site }j} \qquad (12.5)$$

The head loss H_L is assumed to be primarily due to friction.

The friction loss is computed based on 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 (EGL) between sites i and j.

$$\begin{aligned} H_L &= \left(\text{EGL}_i - \text{EGL}_j\right) \\ &= \left[Z + V^2/2g\right]_{\text{site }i} - \left[Z + V^2/2g\right]_{\text{site }j} \end{aligned} \qquad (12.6)$$

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.

12.3.2   Pressure Pipe Networks

The Hazen–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) (that equals the half the pipe radius, r), and the cross-sectional area, A (m^2).

$$V = 0.849CR^{0.63}S^{0.54} \qquad (12.7)$$

$$Q = AV = \pi r^2 V \qquad (12.8)$$

The head loss along a length L (m) of pipe of diameter D (m) containing a flow of Q (m³/s) is defined as

$$H_L = KQ^{1.85}, \qquad (12.9)$$

where K is the pipe coefficient defined by Eq. 12.10.

$$K = [10.66L] / [C^{1.85}D^{4.87}] \qquad (12.10)$$

Another pipe flow equation for head loss is the Darcy–Weisbach equation based on a friction factor f.
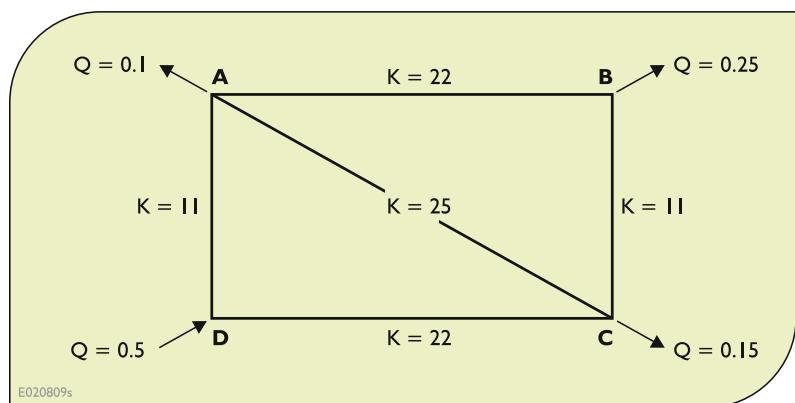
$$H_L = fLV^2/D2g \qquad (12.11)$$

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:

$$Storage_{it} + Q_{it}^{in} - Q_{it}^{out} = Storage_{i,t+1} \quad (12.12)$$

In each section between nodes i and j:

$$H_{Lit} = H_{Ljt} + H_{Lijt}, \qquad (12.13)$$

where the head loss between nodes i and j is H_{Lijt}.

To compute the flows and head losses at each node in Fig. 12.6 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 nonnegative flow variables.

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:

$$0.5 = Q_{DA} + Q_{DC} \qquad (12.14)$$

$$0.1 = Q_{DA} - Q_{AC} + Q_{CA} - Q_{AB} \qquad (12.15)$$

$$0.25 = Q_{AB} + Q_{CB} - Q_{BC} \qquad (12.16)$$

$$0.15 = Q_{DC} + Q_{AC} - Q_{CA} + Q_{BC} - Q_{CB} \qquad (12.17)$$

Continuity of heads at each node requires:

Fig. 12.6 An example of a pipe network, showing the values of K for predicting head losses from Eq. 12.10

$$H_D = H_C + 22 * \left(Q_{DC}^{1.85}\right); \qquad (12.18)$$

$$H_D = H_A + 11 * \left(Q_{DA}^{1.85}\right); \qquad (12.19)$$

$$H_A = H_B + 22 * Q_{AB}^{1.85}; \qquad (12.20)$$

$$H_C = H_A + 25 * \left((Q_{CA} - Q_{AC})^{1.85}\right); \quad (12.21)$$

$$H_C = H_B + 11 * \left((Q_{CB} - Q_{BC})^{1.85}\right) \quad (12.22)$$

Solving these Eqs. 12.14–12.22 simultaneously for the five flow and four head variables yields the flows Q_{ij} from nodes i to nodes j and heads H_i at nodes i listed in Table 12.1.

This solution shown in Table 12.1 assumes that the network is at a constant elevation, has no storage capacity, and there are no minor losses. Losses are usually expressed as a linear function of the velocity head, due to hydraulic structures (such as valves, 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.

12.3.3 Water Quality

Many of the water quality models discussed in Chap. 10 can be used to predict water quality constituent concentrations in open channels and in pressure pipes. The assumption of complete mixing such as at junctions or in short segments of pipe, is made. Reactions among constituents can occur as water travels through the system at predicted velocities. Water-resident times (the ages of waters) in 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

Table 12.1 Flows and heads of the network shown in Fig. 12.6

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

assume combinations of plug flow (advection) along pipe sections and complete mixing within segments of the 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. (See Chap. 10).

Computer programs (e.g., EPANET) exist that can perform simulations of the flows, heads, and water quality behavior 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 applications 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 rechlorination 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.

12.4   Wastewater Collection

12.4.1   Sewer Networks

Flows in urban sewers 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 modeled using one-dimensional flow equations.

When modeling 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 modelers are the entrance and exit losses to the conduit. Typically at each end of the conduit is a manhole. Manholes are storage chambers that provide access (for men and women) to the conduits upstream and downstream. Manholes induce some additional head loss.

Manholes usually cause a major part of the head losses in sewer systems. A manhole head loss represents a combination of the 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.

$$H_L = K\left(V_D^2/2g\right)\left[1 - (A_D/A_U)\right]^2 \qquad (12.23)$$

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, that is, the maximum flow that can occur without surcharging or flooding. Assuming 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.

12.5   Wastewater Treatment

The wastewater generated by residences, businesses, and industries in a community is largely water. Wastewater 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 biodegradable constituents in the wastewater is its biochemical oxygen demand, or BOD_5. BOD_5 is the amount of dissolved oxygen aquatic microorganisms consumed in 5 days as they metabolize (eat) the organic material in the wastewater. Untreated sewage typically has a BOD_5 concentration ranging from 100 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 (Fig. 12.8).

As illustrated in Fig. 12.7, wastewater treatment is a multistage 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.
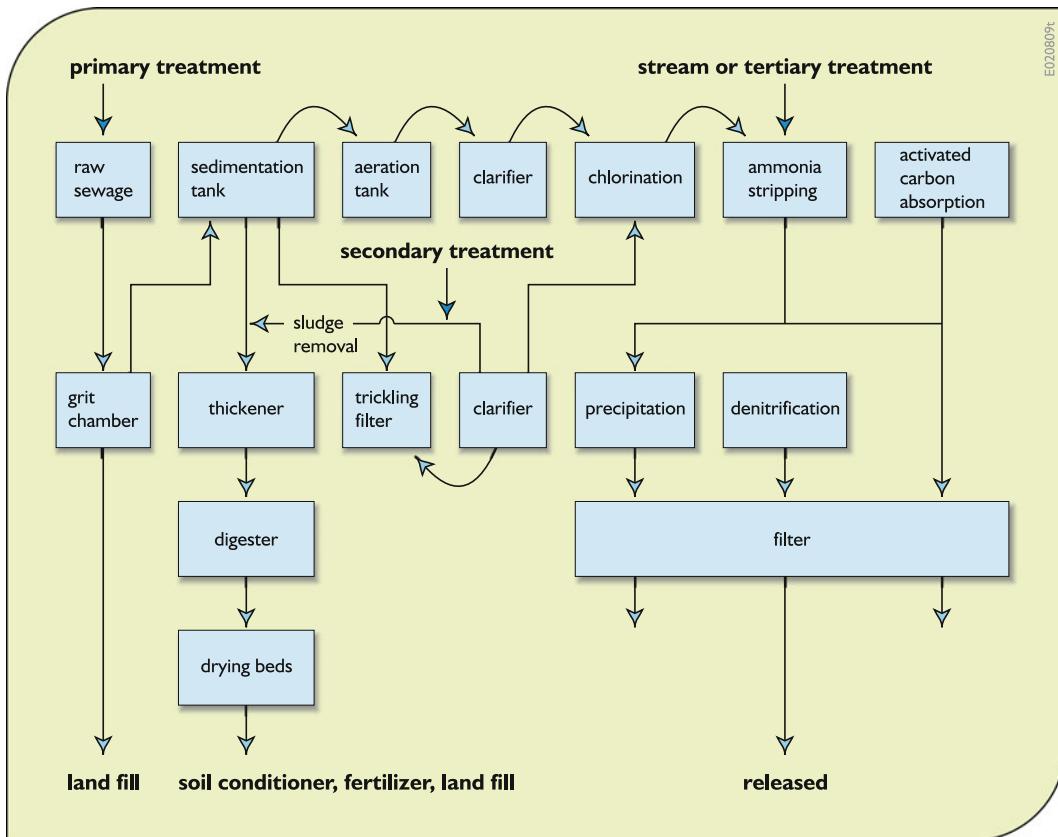
Preliminary treatment removes solid materials (sticks, rags, large particles, sand, gravel, toys, money, or anything people flush down toilets). Equipment such as bar screens, and grit chambers are used to filter the wastewater as it enters a treatment plant. The wastewater then passes on to what is called primary treatment.

Clarifiers and septic tanks are usually used to provide primary treatment. Primary treatment separates suspended solids and greases from wastewater. 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 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.

Secondary treatment is typically a biological treatment process designed to remove dissolved organic matter from wastewater. Sewage microorganisms cultivated 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. The wastewater is spread over the substrate. 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. Activated sludge, extended aeration, oxidation ditch, and sequential batch reactor systems are all examples of suspended film systems. 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.

Fig. 12.7 A typical wastewater treatment plant showing the sequence of processes for removing impurities

Fig. 12.8 Wastewater treatment plant on the shores of Western Lake Superior in Minnesota, USA. The Activated Sludge system is designed to treat 48 million gallons of wastewater per day (mgd) with a peak hydraulic capacity of 160 mgd. (Photo courtesy of Western Lake Superior Sanitary District.)

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 treatment 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 include 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. 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 odors, remove some of the water and reduce volume, decompose some of the organic matter and reduce volume, and kill disease-causing organisms. Following sludge treatment, liquid and cake sludge free of toxic compounds can be spread on fields, returning organic matter and nutrients to the soil.

12.6   Urban Drainage Systems

Urban drainage involves a number of hydraulic and biochemical processes. These typically include:

- Rainfall and surface runoff
- Surface loading and washoff of pollutants
- Stormwater sewer and pipe flow
- Sediment transport
- Separation of solids at structures
- Outfalls

These components or processes are briefly discussed in the following subsections.

12.6.1   Rainfall

Surface runoff of precipitation and the need to collect urban wastewater are the primary reasons for designing and maintaining urban drainage systems. Storms are a major source of flow into the system. Even sanitary sewer systems that are designed 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 in estimating precipitation patterns over space and time, as if more rain gauges were used and as if 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 in 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. These events 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, i.e., 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 analyze system performance under design flow conditions.

12.6.1.1   Time Series Versus Design Storms

Professionals can argue over whether infrastructure design capacities should be based on real rainfall records or synthetic storm events. The

argument in favor of using synthetic storms is that they are easy to use and require only a few events to assess the system design performance. The argument in favor of a time series of real rainfall is that these data typically include a wider range of conditions, and therefore are likely to contain the conditions that are critical on 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.

12.6.1.2 Spatial–Temporal Distributions

Rainfall varies in space as well as in time, and the two effects are related. Short duration storms typically come from small rain cells that have a short life, or that move rapidly over the catchment. As these cells are small (of the order of a kilometer in diameter) there is significant spatial variation in rainfall intensity. Longer duration 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 the use of recorded rainfall data requires some way to account for this spatial and temporal variations. The average rainfall over the catchment in any period of time can be more or less than the measured values at one or more gauge sites. The runoff from a portion of a catchment exposed to a high intensity rainfall will more than the runoff from the same amount of rainfall spread evenly over the entire catchment.

12.6.1.3 Synthetic Rainfall

A convenient way of using rainfall data is to analyze 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.
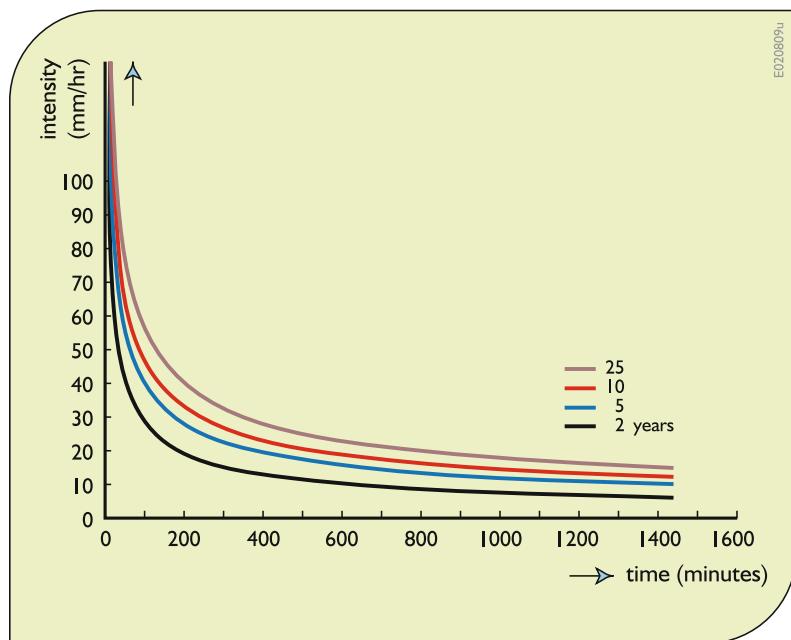
In most of the work on urban drainage and river modeling, the risks of occurrence are expressed not by probabilities but by the inverse of probability, the return period. An event that has a probability of 0.2 of being equaled or exceeded each year has an expected return period of 1/0.2 or 5 years. An event having a probability of 0.5 of being equaled 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 analyzed 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 Fig. 12.9, are derived by analysis of a long set of rainfall records. Intensity–duration–frequency data is commonly available all over the world and therefore it is important to be aware of its method of derivation and ways it can be used for simulation modeling.

Fig. 12.9 Rainfall
intensity–duration–
frequency (return period)
curves

The depth of rainfall is the intensity times its duration integrated over the total storm duration.

12.6.1.4 Design Rainfall

Design rainfall events (hyetographs) for use in simulation models are derived from intensity–duration–frequency data.

The rainfall intensity during an event is not uniform in time, and its variation both in intensity and when the peak intensity occurs during the storm can be characterized by the peakedness of the storm and the skew of the storm (Fig. 12.10).

A design storm is a synthetic storm that has an appropriate peak intensity and storm profile.

12.6.2 Runoff
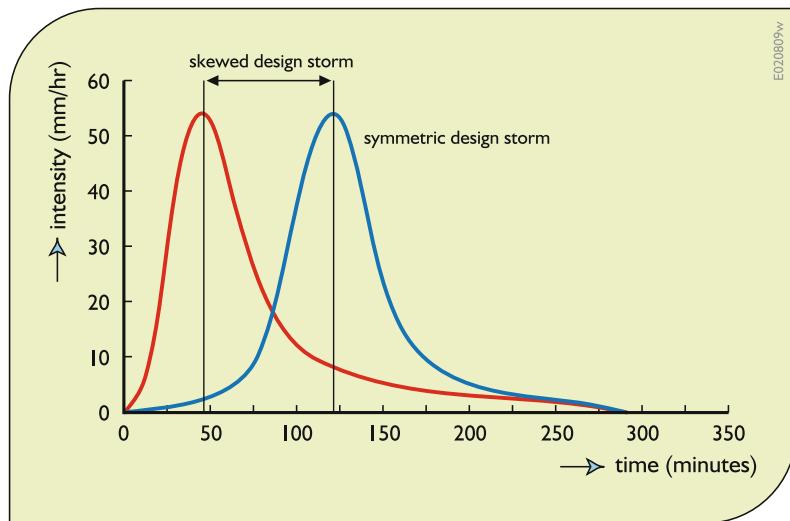
The runoff from rainfall involves a number of processes and events, as illustrated in Fig. 12.11, and can be modeled using various methods. Most of these methods assume an initial loss, a continuing loss, and a remainder contributing to the system runoff.

Most models assume that the first part of a rainfall event goes to initial wetting of surfaces and filling depression storage. The depth assumed to be lost is usually related to the surface type and condition. Rain water 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.
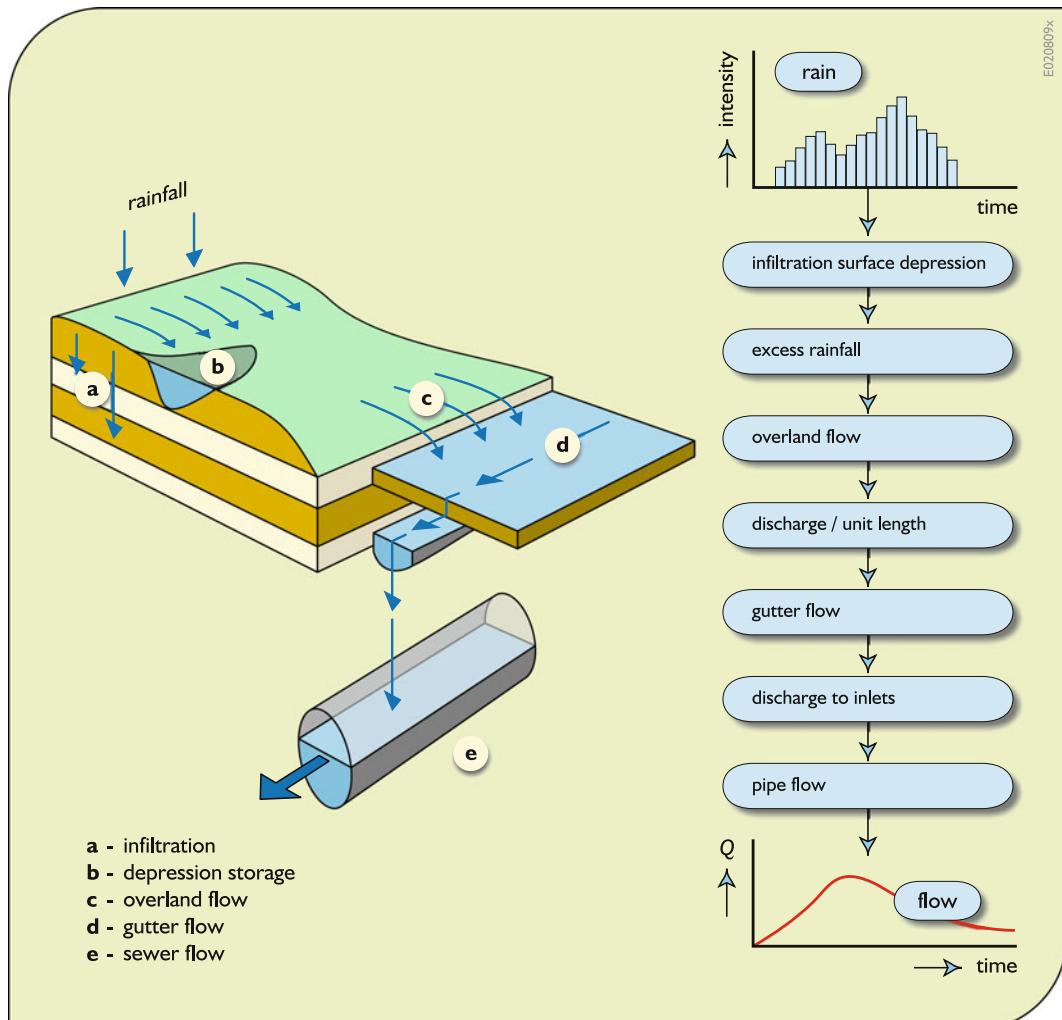
Initial loss depths are defined as the minimum quantity of rainfall causing overland runoff. The initial loss depth of rainfall for catchment surfaces can be estimated as the intercept on the rainfall axis of plots of rainfall verses runoff (Fig. 12.12). The runoff values shown in Fig. 12.12 were obtained for various catchments in the UK (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 subcatchment 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.
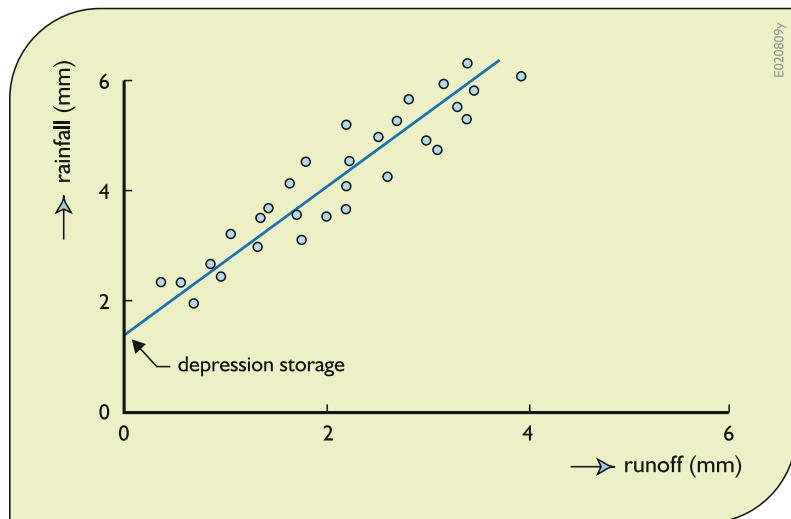
Fig. 12.10 Storm peak skewness profiles

Fig. 12.11 Schematic representation of urban rainfall–runoff processes

Fig. 12.12 Estimation of depression storage based on data from catchments in the UK (Price 2002)

For example, considering 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 evapo-ration 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 vegetation transpiration and gen-eral evaporation are not particularly an issue for single events, but can be during the interevent periods where catchment drying takes place. This is applicable to models where time-series data are used and generated. 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.
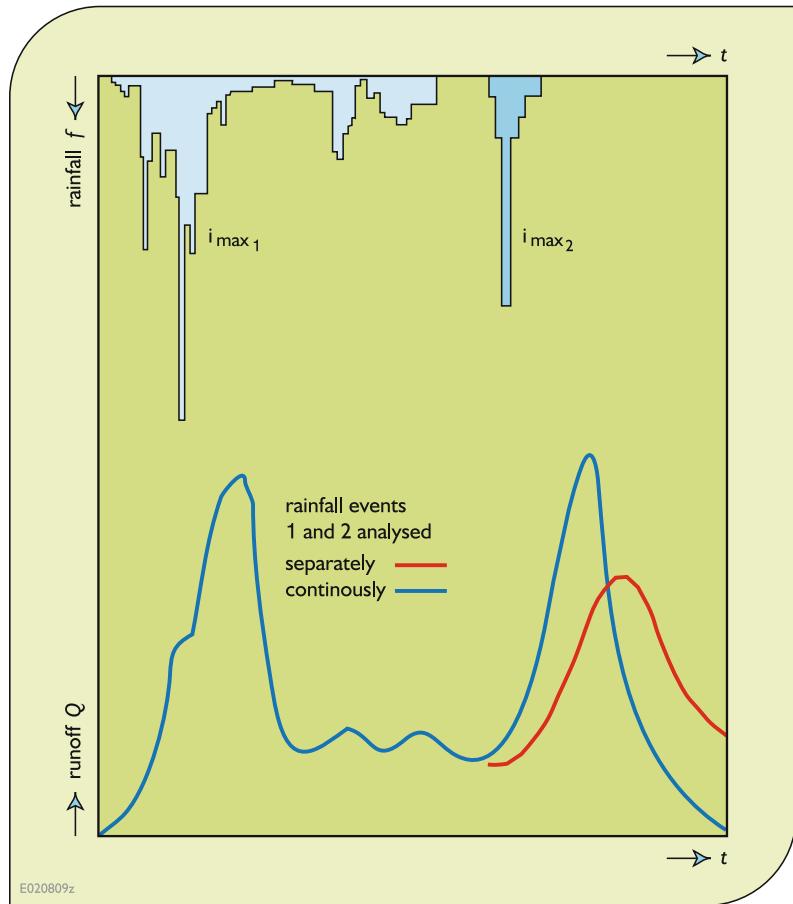
Many models try to account for the wetting of the catchment and the increasing runoff that takes place as wetting increases. The effect of this is shown in Fig. 12.13.

It is impractical to take full account of the variability in urban topography, and surface condition. Impervious (paved) surfaces are often dominant in an urban catchment and the loss of rainfall prior to runoff is usually relatively small.

Runoff routing is the process of passing rainfall across the surface to enter the drainage network. This process results in attenuation and delay. These are modeled using routing tech-niques that generally consider catchment area size, ground slope, and rainfall intensity in determining the flow rate into the network. The topography and surface channels and even upstream parts of the sewer system are usually lumped together into 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.

Various models for rainfall–runoff and rout-ing are available and are used in different parts of the world. Overland runoff on catchment surfaces can be represented by the kinematic wave equa-tion. 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 subcatch-ments. Simpler reservoir-based models, that are less computationally and data demanding, rep-resent the physical processes almost as accurately as the more complex physically based approa-ches (Price 2002). In practice, models applied to catchments typically assume an average or combined behavior of a number of overland flow planes, gutters, and feeder pipes. Therefore, the

Fig. 12.13 Effect of catchment wetness on runoff Q over time t (Price 2002)

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 that the outflow, $Q(t)$ (m³/s), at the catchment outlet is proportional to the volume of stormwater, $S(t)$ (m³), present on the ground surface of that catchment including the nonexplicitly modeled 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 millimeter(s) of rainfall may not contribute to the runoff.

The basic equation for runoff $Q(t)$ (m³/s) at time t is

$$Q(t) = S(t)/K, \qquad (12.24)$$

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 Eq. 12.24, 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 12.2 (Price 2002). Use of these coefficients should be supported either by field observations or by expert judgment.

Table 12.2 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

12.6.2.1   The Horton Infiltration Model

The Horton model describes the increasing runoff from permeable surfaces as 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.

Based on infiltrometer studies on small catchments Horton defined the infiltration rate, f, either on pervious surfaces or on semipervious surfaces, as a function of time, t (hours), the initial infiltration rate, f_o (mm/h), the minimum (limiting or critical) infiltration rate, f_c (mm/h), and an infiltration rate constant, k (1/h).

$$f = f_c + (f_o - f_c)e^{-kt} \qquad (12.25)$$

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 Eq. 12.25 over time defines the cumulative infiltration $F(t)$.
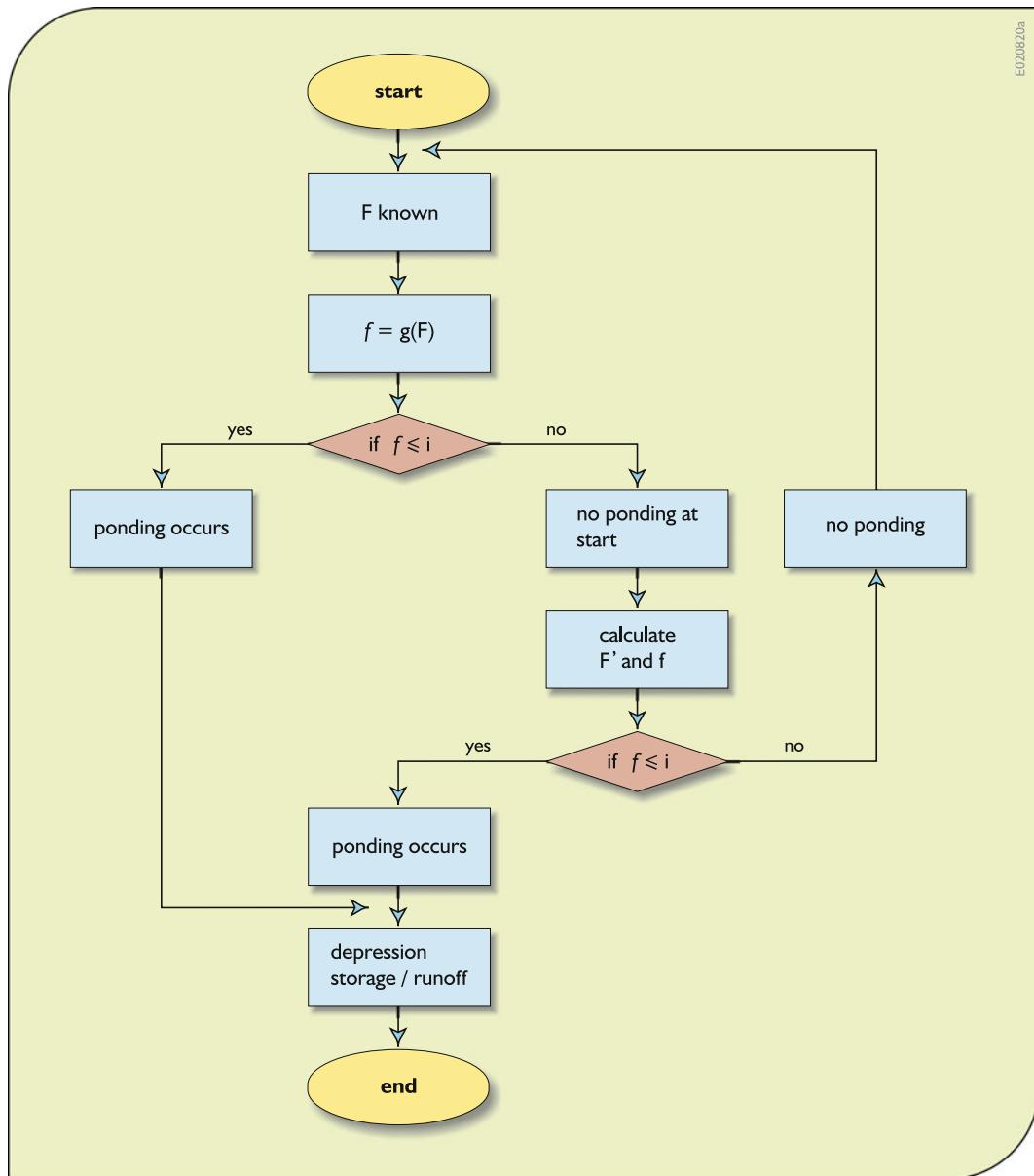
$$F(t) = f_c t + (f_o - f_c)\left(1 - e^{-kt}\right)/k \qquad (12.26)$$

The Horton equation variant as defined in Eq. 12.26 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 Eq. 12.26 may be solved iteratively to find the time it takes to cause ponding, t_p, as a function of F. That time t_p is used in Eq. 12.25 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 Fig. 12.14.

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 12.3, 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.

Fig. 12.14 Flow chart of Horton model infiltration algorithm used in each time step of a simulation model

Table 12.3 Values for Horton's infiltration model for different soil groups as defined by the US Soil Conservation Service

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

12.6.2.2 The US Soil Conservation Method (SCS) Model

The SCS method is a widely used model for predicting runoff from rural catchments, especially in the USA, France, Germany, Australia, and parts of Africa. 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

$$R = I_a + F + Q \quad (12.27)$$

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

$$F/S = Q/(R-I_a), \quad (12.28)$$

when $R > I_a$. These equations combine to give the SCS model

$$Q = (R-I_a)^2/(R-I_a+S) \quad (12.29)$$

This model can be modified for use in continuous simulation models.

Numerical representation of the derivative of Eq. 12.29 can be written as Eq. 12.30 for predicting the runoff, q (mm/Δt), over a time interval Δt given the rainfall r (mm/Δt), in that time interval.

$$q = r(R-I_a)(R-I_a+2S)/(R-I_a+S)^2 \quad (12.30)$$

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

$$I_a = k\,S, \quad (12.31)$$

where $0 < k < 0.2$.

The original SCS approach recommended $k = 0.2$. However, other studies suggest that k values between 0.05 and 0.1 may be more appropriate.

The storage variable, S, itself is 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

$$S = (25400/CN) - 254 \quad (12.32)$$

Curve number values depend on antecedent moisture conditions (AMC) and hydrologic soil

group. The antecedent moisture conditions are divided into three classes, as defined in Table 12.4.

The four hydrologic soil groups are defined in Table 12.5.

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 12.6 lists some of these relevant to urban areas and antecedent moisture condition class AMC II.

Figure 12.15 identifies the CN values for antecedent moisture content (AMC) classes I and III based on class II values.
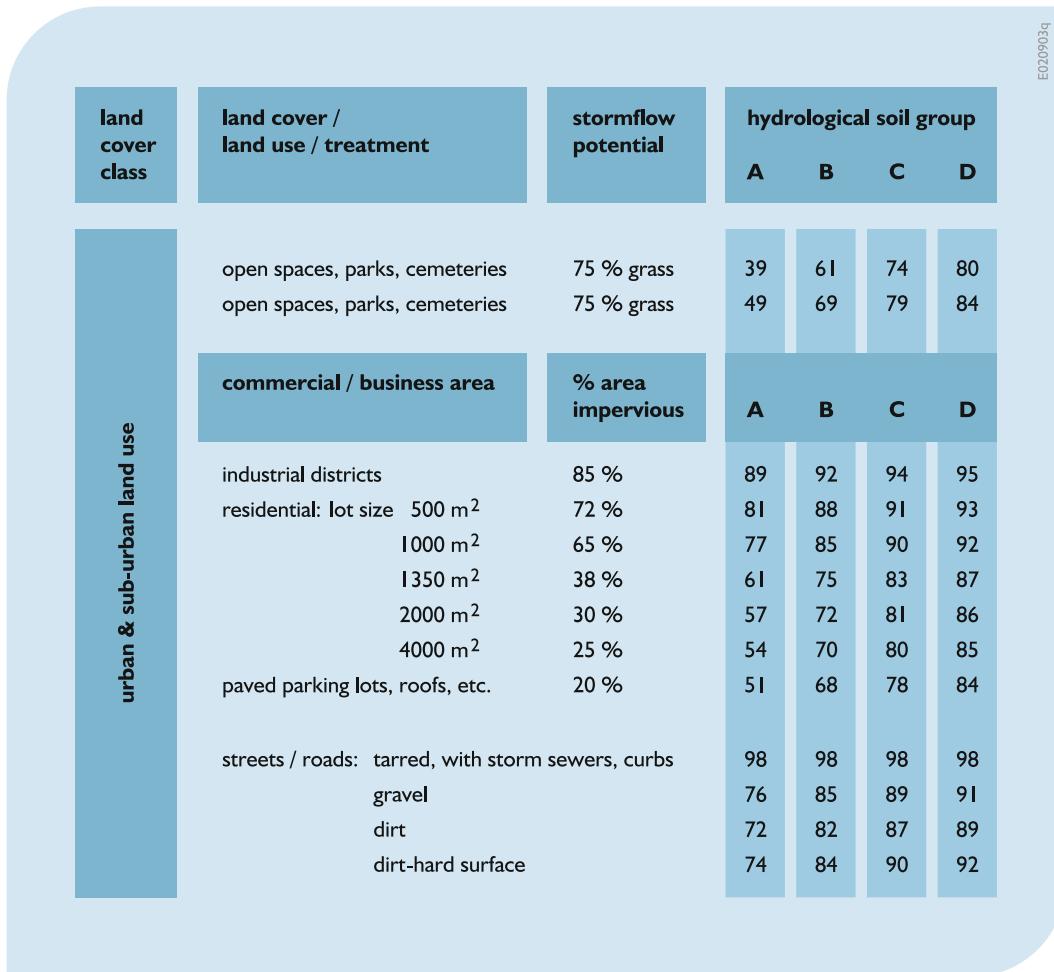
Table 12.4 Antecedent moisture classes (AMC) for determining curve numbers *CN*

	total 5-day antecedent rainfall (mm)	
	season	
AMC	dormant	growing
I	< 12.5	< 35.5
II	12.5 - 28.0	35.5 - 53.5
III	> 28.0	> 53.5

E020903o

Table 12.5 SCS hydrologic soil groups used in Tables 12.3 and 12.6

soil type	definition
A	(low runoff potential) high infiltration rates even when thoroughly wetted. chiefly deep, well to excessively drained sands or gravels. high rate of water transmission.
B	moderate infiltration rates when thoroughly wetted. chiefly moderately deep to deep, moderately-well to well drained soils with moderately-fine to moderately-coarse textures. moderate rate of water transmission.
C	slow infiltration rates when thoroughly wetted. 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) very slow infiltration and transmission rates when thoroughly wetted. chiefly: • clay soils with a high swelling potential • soils with a permanent high water table • soils with a clay pan or clay layer at or near the surface • shallow soils over nearly impervious material

E020903p

Table 12.6 Initial *CN* values for AMC II with various urban land use, cover quality, and hydrologic soil groups (defined in Table 12.5)

land cover class	land cover / land use / 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 m^2	72 %	81	88	91	93
	1000 m^2	65 %	77	85	90	92
	1350 m^2	38 %	61	75	83	87
	2000 m^2	30 %	57	72	81	86
	4000 m^2	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
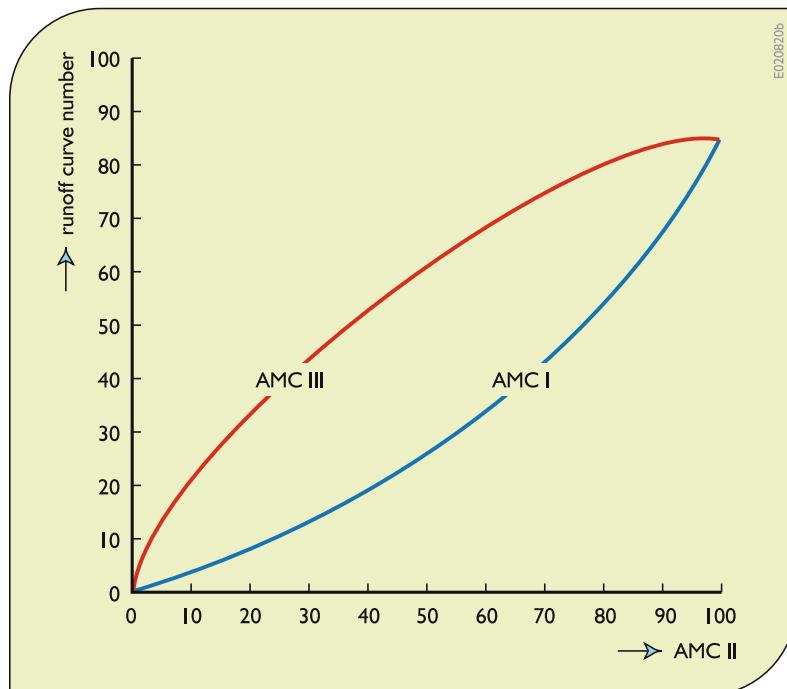
12.6.2.3   The SWMM Rainfall–Runoff Model

The rainfall–runoff element of the popular SWMM model generally assumes 100% runoff from impermeable surfaces and uses Horton or the Green–Ampt model 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 semipervious catchments (Huber and Dickinson 1988; Roesner et al. 1988).

12.6.3   Surface Pollutant Loading and Washoff

The modeling of surface pollutant loading and washoff into sewer 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 vehicular transport emissions. When rainfall takes place some of this material, as dissolved pollutants and fine solids, is washed

Fig. 12.15 Runoff curve
numbers for AMC classes I
and III based on curve
numbers for AMC II

into the stormwater sewers or gullies. During
buildup time many of the pollutants degrade.

Deposition of this material is not homoge-
neous but rather is a function of climate, geog-
raphy, land use, and human activity. The
mechanism of washoff is obviously a function of
location, land use, rainfall intensity, slope, flow
rate, vehicle disturbance, etc. None of these
factors are explicitly modeled in most washoff
and sewer flow models.

Measurements made of pollutant accumula-
tion 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.
Modeling does not change this, it only has the
potential of making those guesses better.

12.6.3.1   Surface Loading

Pollutant loadings and accumulation on the sur-
face of an urban catchment occur during dry
periods between rainstorms.

A common hypothesis for pollutant accumu-
lation 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,
neither over space nor 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 Chaps. 6 and 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 accumu-
lated mass M_P (kg/ha). Assuming a proportion-
ality constant (decay rate constant) of k_P (1/day),
the rate of change in the accumulated mass M_P
over time t is

$$dM_P/dt = m_P - k_P M_P \qquad (12.33)$$

As the number of days during the dry period gets 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 Eq. 12.33 over the duration Δt (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.

$$M_P(\Delta t) = M_P(0)\,e^{-kp\Delta t} + \left[m_P\left(1 - e^{-kp\Delta t}\right)/k_P\right],$$
$$(12.34)$$

where $M_P(0)$ is the initial mass of pollutant P on the catchment surface at the beginning of the dry period, i.e., at the end of the previous rainstorm.
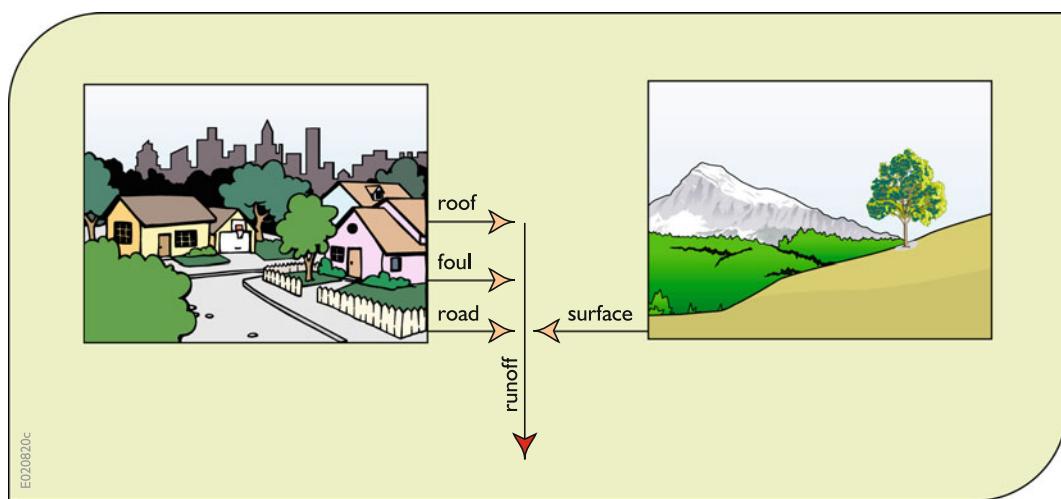
Sediments (that become suspended solids in the runoff) are among the pollutants accumulating on the surface of urban catchments. They are important by themselves, but also because some of the other pollutants that accumulate become attached to these sediments. Sediments are typically defined by their medium diameter size value (d_{50}). Normally a minimum of two sediment fractions are modeled, 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.

12.6.3.2   Surface Washoff

Pollutants in the washoff can be dissolved in water, or they can be attached to the sediments. Many models of the transport of dissolved and particulate pollutants through a sewerage system assume each pollutant is conservative—it does not degrade with 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 sewer 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 rainfall and by erosion from runoff flowing across the surface. Figure 12.16 shows schematically some sources of pollution in the washoff model.

Fig. 12.16 Some sources of pollutants in the washoff to sewer systems from land surfaces

The rate of pollutant washoff is dependent on an erosion coefficient, α_P, and the quantity, M_P, of available pollutant, P. As the storm event proceeds and pollutants 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 that runoff. 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:

$$f_t^P = \alpha_P R_t / (1 + \alpha_P R_t) \qquad (12.35)$$

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 α_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 based on 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 usually are 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,

$$M_{Pt+1} = \left(1 - f_t^P\right) M_{Pt} e^{-kP} + m_P \qquad (12.36)$$

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 modeled 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.

12.6.3.3   Stormwater Sewer and Pipe Flow

Flows in pipes and sewers have been analyzed 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 manholes. Difficulties in obtaining convergence occurs at pipes with steep to flat transitions, dry pipes, etc., and therefore additional features and checks are needed to achieve satisfactory model results.

12.6.3.4   Sediment Transport

Pollutant transport modeling 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 bioreactor is not included in most simulation models. Most models assume pollutants are conservative during the time in residence in the drainage system before being discharged into a water body. All these processes that take place in transient are generally either ignored or approximated using a range of assumptions.

12.6.3.5   Structures and Special Flow Characteristics

Manholes, valves, pipes, pumping stations, overflow weirs, etc., that impact 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 2- or 3-dimensional models. However, the ever-increasing power of computers is making higher dimensional fluid dynamic analyses increasingly available to practicing engineers. The biggest limitation may be more related to data and calibration than to computer models and costs.

12.6.4   Water Quality Impacts

12.6.4.1   Slime

Slime can build up on the perimeter 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-full 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.

12.6.4.2   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 flowing part full and when considerable energy is lost as a result of the generation of surface disturbances. In practice the lower roughness values are used as flow states of interest are usually extreme events and therefore sewers are operating in surcharge.

12.6.4.3   Pollution Impact on the Environment

The effects of combined sewer outflows (CSOs) 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 CSOs are one of the major causes of poor river water quality. Standards and performance criteria specifically for intermittent discharges are therefore needed to reduce the pollutants in CSOs.

Drainage discharges that affect water quality include:

(1) 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),
(2) substances that physically hinder reoxygenation at the water surface, such as oils,
(3) discharges containing toxic compounds, including ammonia, pesticides, and some industrial effluents, and
(4) discharges that are high in suspended solids and thus inhibit biological activity by excluding light from 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 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 odors. It can also be a direct hazard to health, particularly in tropical regions where waterborne diseases such as cholera and typhoid prevail.
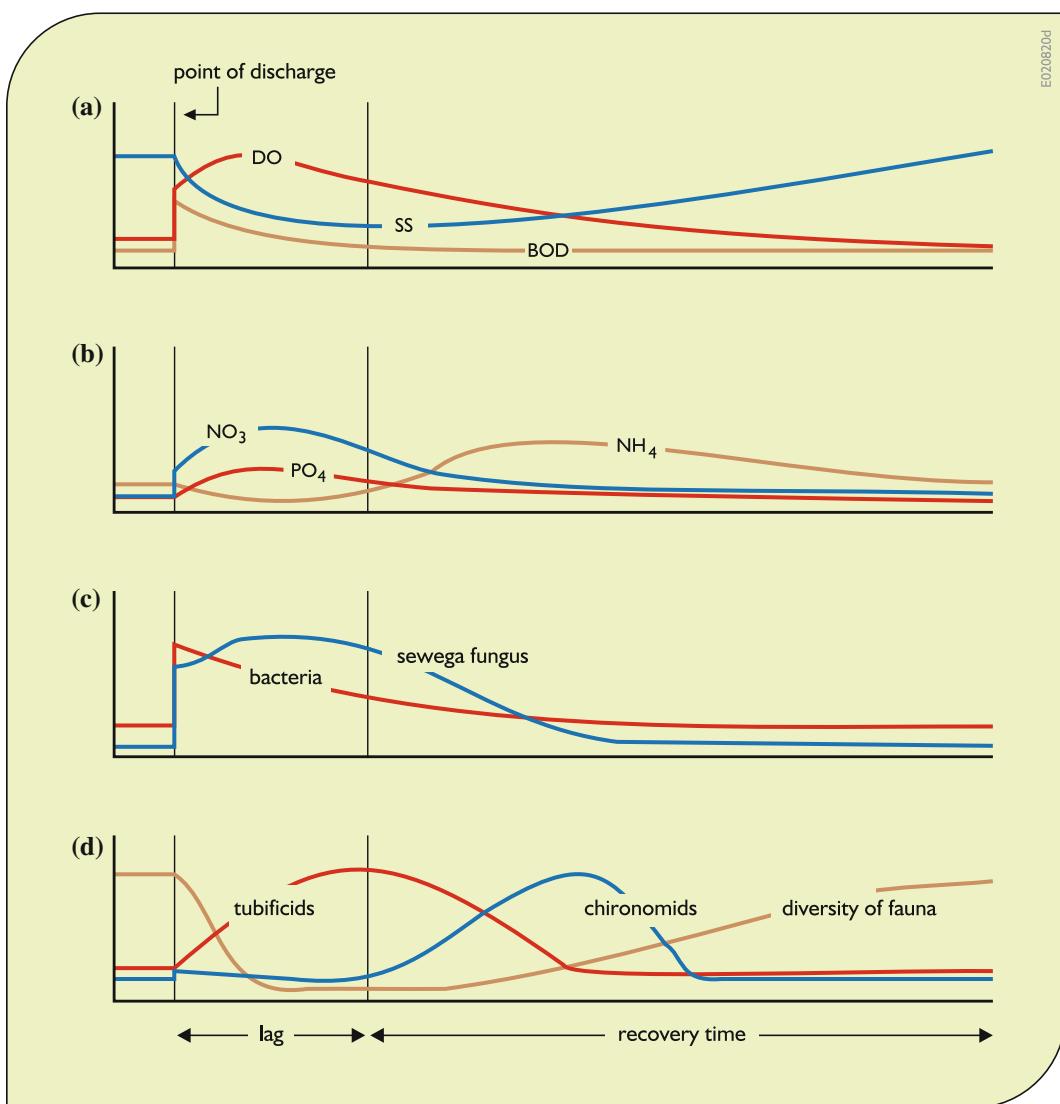
The aim of good drainage design, with respect to pollution, is to balance the effects of

continuous and intermittent discharges against the assimilation capacity of the water, so as to achieve in a cost and socially effective way the desired quality of the receiving water.

Figure 12.17 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 they 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 Fig. 12.17 the *BOD* in diagram "a" rises or stays constant after release despite some of the

Fig. 12.17 Pollution impact along a waterway downstream from its discharge

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 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 Fig. 12.17 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 due to the very low tolerance many aquatic organisms, particularly fish, have to the chemical. The ammonium concentration rises if the conditions are anaerobic and will then decline once aerobic conditions return and the ammonium ions are oxidized to nitrates.

Diagrams "c" and "d" show the effect of combined sewer overflows (CSOs) 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 lead 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 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 is 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 breakdown 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 liter of water. These organisms take time to fully digest and stabilize the waste. The rate at which they do so depends on the temperature of, and the quantity of these organisms available in, the water at the start.

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 5 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 12.7 and 12.8 provide some general ranges of pollutant concentrations in CSOs from urban catchments.

Water quality models for urban drainage are similar, or simpler versions of water quality models for other water bodies (as discussed in

Table 12.7 Typical quality of domestic sewage

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

Table 12.8 Pollutant concentrations (mg/l) in urban runoff

constituent	highway runoff	residential area	commercial area	industrial area
suspended solids mg/l	28 - 1178	112 - 1104	230 - 1894	34 - 374
BOD5 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

Chap. 10). Often just a simple mixing and dilution model will be sufficient to predict the concentration of pollutants at any point in a sewer 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 Chap. 10 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 noncontinuous 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.

12.6.4.4   Bacteriological and Pathogenic Factors

The modeling 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 24 h 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. This indicator is commonly used throughout the world to test water samples for fecal contamination.

Because of the problems in measuring microbiological parameters involved in CSOs, 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.

12.6.4.5   Oil and Toxic Contaminants

Oils are typically discharged into sewers by people, 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 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 CSOs:

- 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 that 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 sulfide can also be significant.

12.6.4.6   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 also can 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.

12.6.5   Green Urban Infrastructure

Many urban areas are being recognized for successfully using natural ecosystems to reduce the costs of providing clean drinking water and managing stormwater. Moreover, this "green" infrastructure provides many quality-of-life benefits, by improving air quality, increasing shading, contributing to higher property values, and enhancing streetscapes. New York City in the US, for example, estimates that such efforts have saved ratepayers billions of dollars—by eliminating the need for construction of hard "gray" infrastructure such as storm sewers and filtration plants—while preserving large tracts of natural areas. The department's *Green Infrastructure Plan* lays out how the city will improve the water quality in New York Harbor by capturing and retaining stormwater runoff before it enters the sewer system using streetside swales, tree pits, and blue and green rooftop detention techniques to absorb and retain stormwater (Fig. 12.18). This hybrid approach reduces combined sewer overflows by 12 billion gallons a year—over 2 billion gallons a year more than the current all-gray strategy—while saving New Yorkers $2.4 billion (NYCDEP 2012).

12.7   Urban Water System Modeling

Optimization and simulation models are becoming increasingly available and used to analyze a variety of design and operation problems involving urban water systems. Many are incorporated within graphics user interfaces that facilitate the use of the models and the understanding and further analyses of their results.

12.7.1   Optimization

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 nonlinearity of especially urban water distribution networks.

Fig. 12.18 Examples of "*green*" and "*blue*" (*upper center*) rooftop infrastructures for reducing urban storm runoff and at the same time enhancing the attractiveness of the urban environment. Some go to extremes

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 (Ormsbee et al. 1986; Boulos et al. 1990). 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 et al. (1991), Datta et al. (1994), Reddy et al. (1996), Pudar et al. (1992), and Liggett et al. (1994) to solve various steady-state and transient model calibration problems (Datta et al. 1994; Savic et al. 1995; Greco et al. 1999; Vitkovsky et al. 2000).

Evolutionary search algorithms, discussed in Chap. 5, are now commonly used for the design and calibration of various highly nonlinear 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 et al. 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, 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.

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 experiencing multiple increasing demands can be much more challenging.

Consider, for example, developing a master plan for the next 20 years divided into four 5-year construction periods. The obvious way to model this problem is to include the system design variables for each of the next four 5-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.

Dandy et al. (2002) have developed and applied two alternative modeling 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 20-year planning period. A series of subproblems are then optimized, one for each intermediate planning stage,

to identify when each facility that is to be built should be built. For these subproblems, 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 latter buildup method finds the optimal solution for the first planning stage that 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 buildup 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 judgment, can be the basis of 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 with updated information can be performed again to obtain a better estimate of what the next period's decisions should be.

12.7.2 Simulation

Dynamic simulation models are increasingly replacing steady-state models for analyzing water quantity, pressure and water quality in distribution and collection networks. Dynamic models provide estimates of the time-variant behavior 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 its nature is stochastic. Changes in water use, control responses and dispatch of sources, and random storms over different parts of the catchment all can affect flow quantities, the flow 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 (see Box 12.1).

The simulation of water quantities and qualities in urban catchments serve 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 analyze 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 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).

12.8 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 water is used, the now wastewater needs to be collected and transported to where it can be treated and either reused or discharged back into the environment. Overlaying 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 for predicting the impacts of alternative urban water treatment, distribution, and collection system designs and operating, maintenance and repair policies.

Modeling 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 modelers 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 at the end of this chapter.

References

Al-Zahrani, M. A., & Moied, K. (2001). *Locating water quality monitoring stations in water distribution systems*. ASCE, Orlando, Florida: Proceedings of the World Water and Environmental Resources Congress.

Aral, M. M., Maslia, M. L., Ulirsch, G. V., & 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, 51*(4), 300–309.

Bedient, P. B., & Huber, W. C. (1992). Hydrology and floodplain analysis, 2nd ed. Reading, MA: Addison-Westley Pub. Co.

Boulos, P. F., & Wood, D. J. (1990). Explicit calculation of pipe-network parameters. *Journal of Hydraulic Engineering, ASCE, 116*(11), 1329–1344.

Chin, D. A. (2000). *Water-resources engineering*. Upper Saddle River, NJ: Prentice Hall.

Dandy G., Kolokas, L., Frey, J., Gransbury, J., Duncker, A., & Murphy, L. (2002). Optimal staging of capital works for large water distribution systems, In *Proceedings, ASCE Environmental and Water Resources Planning Symposium, Roanoke, Va.*

Dandy, G. C., & Connarty, M. (1995). Use of genetic algorithms for project sequencing, integrated water resources planning for the 21st century. In *Proceedings of the 22nd Annual Conference, Water Resources Planning and Management Division* (pp. 540–543). Cambridge, Mass: ASCE.

Dandy, G. C., & Engelhardt, M. (2001). The optimal scheduling of water main replacement using genetic algorithms. *Journal of Water Resources Planning and Management, ASCE, 127*(4), 214–223.

Dandy, G. C., Simpson, A. R., & Murphy, L. J. (1996). An improved genetic algorithm for pipe network optimization. *Water Resources Research, 32*(2), 449–458.

Datta, R. S. N., & Sridharan, K. (1994). Parameter estimation in water-distribution systems by least squares. *Journal of Water Resources Planning and Management, ASCE, 120*(4), 405–422.

Department of Environmental Protection. (2012). NYC 2012 Green Infrastructure Annual Report. http://www.nyc.gov/html/dep/pdf/green_infrastructure/gi_annual_report_2013.pdf

Greco, M., & Del Guidice, G. (1999). New approach to water distribution network calibration. *Journal of Hydraulic Engineering, ASCE, 125*(8), 849–854.

Harding, B. L., & Walski, T. M. (2002). Long time-series simulation of water quality in distribution systems. *Proceedings, ASCE Environmental and Water Resources Planning Symposium, Roanoke, Va.*

Huber, W. C., & Dickinson, R. E. (1988). Storm water management model, Version 4, User's Manual. EPA/600/3-88/001a (NTIS PB88-236641/AS), U.S. EPA, Athens, GA, 30605. 595 pp. Also see http://www.epa.gov/ednnrmrl/swmm/.

Lansey, K. E., & Basnet, C. (1991). Parameter estimation for water distribution networks. *Journal of Water Resources Planning and Management, ASCE, 117*(1), 126–144.

Liggett, J. A., & Chen, L. C. (1994). Inverse transient analysis in pipe networks. *Journal of Hydraulic Engineering, ASCE, 120*(8), 934–955.

Mays, L. W. (Ed.). (2000). *Water distribution systems handbook*. New York: McGraw-Hill.

Mays, L. W. (2001). *Water resources engineering*. New York: Wiley.

Murphy, L. J., Simpson, A. R., Dandy, G. C., Frey, J. P., & Farrill, T. W. (1996). Genetic algorithm optimization of the Fort Collins-Loveland water distribution systems, *Specialty Conference on Computers in the Water Industry* (pp. 181–185). Chicago, Illinois: AWWA.

Ormsbee, L. E., & Wood, D. J. (1986). Explicit pipe network calibration. *Journal of Water Resources Planning and Management, 112*(2), 166–182.

Price, R. K. (2002). *Urban drainage modeling, course notes*. Delft, NL: International Institute for Infrastructural, Hydraulic and Environmental Engineering.

Pudar, R. S., & Liggett, J. A. (1992). Leaks in pipe networks. *Journal of Hydraulic Engineering, ASCE, 118*(7), 1031–1046.

Rahal, C. M., Sterling, M. J. H., & Coulbeck, B. (1980). Parameter tuning for simulation models of water distribution networks. *Proceedings of Institution of Civil Engineers*, Part 2, 69, 751–762.

Reddy, P. V. N., Sridharan, K., & Rao, P. V. (1996). WLS method for parameter estimation in water distribution networks. *Journal of Water Resources Planning and management, ASCE, 122*(3), 157–164.

Rodenbeck, S. E., & 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, 2*(2), 53–61.

Roesner, L. A., Aldrich, J. A., & 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. 203 pp.

Savic, D. A., & Walters, G. A. (1995). *Genetic algorithm techniques for calibrating network models, Report No. 95/12*, Centre for Systems and Control Engineering, University of Exeter, p. 41.

Savic, D. A., & Walters, G. (1997). Genetic algorithms for least-cost design of water distribution networks. *Journal of Water Resources Planning and Management, ASCE, 123*(2), 67–77.

Savic, D. A., Walters, G. A., Randall Smith, M., & Atkinson, R. M. (2000). Cost savings on large water distribution systems: Design through GA optimisation. *Proceedings, Joint Conference on Water Resources Engineering and Water Resources Planning and Management, ASCE, Minneapolis, Minnesota*.

Shamir, U. (1974). Optimal design and operation of water distribution systems. *Water Resources Research, 10*(1), 27–36.

Simpson, A. R., Dandy, G. C., & Murphy, L. J. (1994). Genetic algorithms compared to other techniques for pipe optimization. *Journal of Water Resources Planning and Management, ASCE, 120*(4), 423–443.

Simpson, A. R., Sutton, D. C., Keane, D. S., & Sheriff, S. J. (1999). Optimal control of pumping at a water filtration plant using genetic algorithms. In D. A. Savic & G. A. Walters (Eds.), Water industry systems: Modelling and optimization applications (Vol. 2, pp. 407–415). Baldock, England: Research Studies Press Ltd.

Tucciarelli, T., Criminisi, A., & Termini, D. (1999). Leak analysis in pipeline systems by means of optimal valve regulation. *Journal of Hydraulic Engineering, ASCE, 125*(3), 277–285.

Vitkovsky, J. P., & Simpson, A. R. (1997). *Calibration and leak detection in pipe networks using inverse transient analysis and genetic algorithms, Report No. R 157*, Department of Civil and Environmental Engineering, University of Adelaide, p. 97.

Vitkovsky, J. P., Simpson, A. R., & Lambert, M. F. (2000). Leak detection and calibration using transients and genetic algorithms. *Journal of Water Resources Planning and Management, ASCE, 126*(4), 262–265.

Walski, T. M. (1983). Technique for calibrating network models. *Journal of Water Resources and Planning Management, ASCE, 109*(4), 360–372.

Webler, T., & Brown, I. S. (1993). Exposure to Tetrachloroethylene via contaminated drinking water pipes in Massachusetts: A predictive model. *Archives of Environmental Health, 48*(5), 293–297.

Additional References (Further Reading)

Alegre, H., & Coelho, S. T. (2012). Infrastructure asset management of urban water systems, chapter 3. In A. Ostfeld (Ed.), Water supply system analysis—selected topics. doi:10.5772/52377, ISBN 978-953-51-0889-4

Axworthy, D. H., & Karney, B. W. (1996). Modeling low velocity/high dispersion flow in water distribution systems. *Journal of Water Resources Planning and Management, ASCE, 122*(3), 218–221.

Bahri, A. (2012). *Integrated urban water management, TEC background paper 16 global water partnership SE-111 51*. Stockholm, Sweden: Printed by Elanders.

Chaudry, M.H., & Islam, M. R. (1995). Water quality modeling in pipe networks. In E. Cabreran & A. F. Vela (Eds.), *Improving efficiency and reliability in water distribution systems*. Dordrecht: Kluwer Academic Publishers.

Clark, R. M., Males, R., & Stevie, R. G. (1984). Water supply simulation model volumes I, II and III., U.S. EPA, Cincinnati, OH.

Clark, R. M., Grayman, W. M., Males, R. M., & Coyle, J. M. (1986). *Predicting water quality in distribution systems", proceedings, distribution system symposium, AWWA, September 7–10* (pp. 155–180). Minnesota: Minnesota.

EPA. (1992). CEAM Systems Development Life Cycle Methodology (SDLCM) Statement of Policy, Standards, and Guidelines—Version 1.00. U.S. EPA, Athens, GA, 30605.

Feyen, J., Shannon, K., & Neville, M. (2008). *Water and urban development paradigms: Towards an integration of engineering, design and management approaches*. Boca Raton: CRC Press.

Grayman, W. M., Clark, R. M., & Males, R. M. (1988). Modeling distribution system water quality: Dynamic approach. *Journal of Water Resources Planning and Management, ASCE, 114*(3), 295–312.

Grayman, W. M., Loucks, D. P., & Saito, L. (Eds.). (2012). *Toward a sustainable water future: Visions for 2050*. Reston, VA: ASCE Press.

Hering, J. G., Waite, T. D., Luthy, R. G., Drewes, J. E., & Sedlak, D. L. (2013). A changing framework for urban water systems environment. *Science Technology, 47*, 10721–10726. dx.doi.org/10.1021/es4007096

Hutton, C. J., Vamvakeridou-Lyroudia, L. S., Kapelan, Z., & Savic, D. A. (UNEXE). (2011). Uncertainty quantification and reduction in urban water systems (UWS) modelling: Evaluation Report D3.6.1 prepared 2011.005, A framework for integrated modeling of urban systems.

Islam, M. R., & Chaudry, M. H. (1998). Modeling of constituent transport in unsteady flows in pipe networks. *Journal of Hydraulic Engineering, ASCE, 124*(11), 1115–1124.

Medows, M. E., Walski, T. M., et al. (Eds.). (1997). *Computer applications in hydraulic engineering*. Waterbury, CT: Haestad Press, Haestad Methods Inc.

Morley, M. S., Atkinson, R. M., Savic, D. A., & Walters, G. A. (2001). GAnet: Genetic algorithm platform for pipe network optimisation. *Advances in Engineering Software, 32*(6), 467–475.

Grimm, N. B., Grove, J. M., Pickett, S. T. A., & Redman, C. L. (2000). Integrated approaches to long-term studies of urban ecological systems. *BioScience, 50*(7), 571–584.

Ormsbee, L. E. (1989). Implicit network calibration. *Journal of Water Resources Planning and Management, ASCE, 115*(2), 243–257.

Ostfeld, A. (ed). (2012). Water supply system analysis—selected topics, ISBN 978-953-51-0889-4, 158 pages, Publisher: InTech, doi:10.5772/2882

Pickett, S. T. A., Cadenasso, M. L., Grove, J. M., Nilon, C. H., Pouyat, R. V., Zipperer, W. C., & Costanza, R. (2001). Linking terrestrial ecological, physical, and socioeconomic components of metropolitan areas, urban ecological systems. *Annual Review of Ecology System, 32*, 127–57.

Rossman, L. A., Boulos, P. F., & Altman, T. (1993). Discrete volume-element method for network water quality models. *Journal of Water Resources Planning and Management, ASCE, 119*(5), 505–517.

Rossman, L. A., Clark, R. M., & Grayman, W. M. (1994). Modeling chlorine residuals in drinking-water distribution systems. *Journal of Environmental Engineering, ASCE, 120*(4), 803–820.

Shamir, U., & Howard, C. D. D. (1968). Water distribution systems analysis. *Journal of the Hydraulic Division, ASCE, 94*(1), 219–234.

Tang, K., Karney, B., Pendlebury, M., & Zhang, F. (1999). Inverse transient calibration of water distribution systems using genetic algorithms. In D. A. Savic & G. A. Walters (Eds.), *Proceedings of Water Industry Systems: Modelling and Optimisation Applications* (Vol. 1, pp. 317–326). Exeter, UK.

Todini, E. (1999). Using a Kalman Filter approach for looped water distribution network calibration. In D. A. Savic & G. A. Walters (Eds.), *Proceedings of Water Industry Systems: Modelling and Optimisation Applications* (Vol. 1, pp. 327–336), Exeter, UK.

United Nations World Water Development Report No. 4. (2012). *World Water Assessment Programme*. Paris: UNESCO.

Viessman, W. Jr., & Welty, C. (1985). *Water Manage-ment Technology and Institutions*. New York: Harper & Row, Pub.

Exercises

12.1 Define the components of the infrastruc-ture needed to bring water into your home and then collect the wastewater and treat it prior to discharging it back into a receiv-ing water body. Draw a schematic of such a system and show how it can be modeled to determine the best design variable val-ues. Define the data needed to model such a system and then make up values of the needed parameters and solve the model of the system.

12.2 Compare the curve number approach to the use of Manning's equation to estimate urban runoff quantities. Then define how you would predict quality and sediment runoff as well.

12.3 Develop a simple model for predicting the runoff of water, sediment, and several chemicals from a 10-ha urban watershed in the northeastern United States during August 1976. Recorded precipitation was as follows:

Day	1	6	7	8	9	10	13	14	15	26	29
R_1(cm)	1.8	0.7	2.6	2.9	0.1	0.3	2.9	0.1	1.4	3.7	0.8

Solids (sediment) buildup on the watershed at the rate of 50 kg/ha-day, and chemical concen-trations in the solids are 100 mg/kg. Assume that each runoff event washes the watershed surface clean. Assume also that there is no initial sediment buildup on August. The watershed is 30% imper-vious. For each storm use your model to compute:

(a) Runoff in cm and m^3.
(b) Sediment loss (kg).
(c) Chemical loss (g), in dissolved and solid-phase form for chemicals with three different adsorption coefficients, $k = 5$, 100, 1000.

12.4 There exists a modest-sized urban subdi-vision of 100 ha containing 2000 people. Land uses are 60% single-family residen-tial, 10% commercial, and 30% undevel-oped. An evaluation of the effects of street-cleaning practices on nutrient los-ses in runoff is required for this catchment.

This evaluation is to be based on the 7-month precipitation record given below. Present the results of the simulations as 7-month PO_4 and N losses as functions of street-cleaning interval and efficiency (i.e., show these losses for ranges of intervals and efficiencies). Assume a runoff threshold for washoff of $Q_o = 0.5$ cm.

Precipitation (cm)							
Day	A	M	J	J	A	S	O
1			0.6		1.4	0.7	1.9
2	1.1	0.4				0.5	
3						0.1	0.1
4	0.1	1.5					
5	0.1	0.9		1.4	1.9	0.3	
6	0.1		1.4		1.0		0.5
7			0.1	1.1	0.7		
S	0.1	0.1		0.7			0.4
9		0.6	1.6	0.1			1.5
10					0.1		
11							
12				02	02		02
13	0.1			0.2		1.5	
14			02		0.5	3.5	0.8
15							1.0
16						4.3	2.8
17			0.7	0.8	0.5	0.8	1.9
18		0.5	0.4	0.1		0.8	0.1
19			0.4			0.4	0.9
20			0.7	03		2.3	
21						0.3	
22	0.1			0.1	0.4		
23	2.0						
24	3.2	0.1			02	4.7	
25	0.1		0.6			2.8	

Precipitation (cm)						
26				3.0		1.6
27						
28			0.3			0.1
29		0 2	1.1			
30		0.1			0.6	
31					0.2	

12.5 Managing the quantity and quality of stormwater runoff is a common urban problem. Discuss the factors to be considered when planning storm sewer networks and detention basins, and how might simulation and/or optimization methods be used to help do this.

12.6 Multiple connected pipeline networks are commonly make up urban water distribution systems. Define a simple pipeline network and develop an optimization model for finding the flows and heads in the network needed to provide required flow discharges and pressures at various nodes or junctions of the network. Discuss some of the complicating issues associated with the design of such networks that your model may not be considering.

12.7 Consider a wastewater treatment plant and associated effluent detention pond designed to release treated and stored effluent into a stream so as to adapt to varying effluent concentration standards associated with varying stream assimilative capacities as its flow and water temperature and existing pollutant loads vary over time. Develop a model for estimating the treatment plant efficiency for BOD removal, and the size of the detention basin, needed to meet the varying effluent standards of the receiving stream, at a minimum cost. Define the data you will need to do this, and all model parameters. Why might this proposed adaptive treatment scheme not be very practical?

12.8 Urban Infrastructural Asset Management is increasingly becoming a key topic in the move toward increased sustainability of water supply and wastewater systems. List some characteristics of sustainable infrastructural systems.

Project Planning: Putting It All Together

<div style="text-align:right">**13**</div>

Water resources planning and management issues are rarely simple. Projects focused on addressing and finding solutions to these issues are also rarely simple. These projects too need to be planned and executed in ways that will maximize their likelihood of success, i.e., will lead to useful results. When decision-makers and other stakeholders disagree over what they want, and what they consider useful and helpful, the challenge facing project planners and managers is even more challenging. This chapter offers some suggestions on project planning and management. These suggestions reflect years of experiences the writers and their institutions, have had planning and participating in various water resources development projects, at various scales, in many river basins and watersheds throughout much of the world.

Each water resources system is unique, and the specific application of any planning and analysis approach needs address the particular issues of concern as well as adapt to the political environment in which decisions are made. What is important in all cases is that such planning and analyses activities are comprehensive, systematic and transparent, and are performed in full and constant collaboration with the region's planners, decision-makers, and the interested and affected public.

13.1 Water Management Challenges

Managing water is important. The effectiveness of strategies for dealing with water availability, quality, and variability is a major determinant of the survival of species, the functioning and resilience of ecosystems, the vitality of societies, and the strength of economies. Humans have been managing water and adapting to surpluses and shortfalls since the dawn of civilization, and especially since the early origins of agriculture. There is evidence across the globe of thousands of years of dam building and canal construction to direct water toward crops of various kinds. Though the tools and infrastructure water managers can use today are dramatically more sophisticated than those used in the past and the scale on which water managers work is much larger in almost all cases, the activities are still very much the same: managing floods and droughts through harvesting and storing water above or underground, delivering water across long distances through pipelines and canals, treating, distributing water supplies to where they are needed, collecting, and treating the resulting wastewaters all designed to meet a variety of economic, public health, environmental, and social objectives.

In regions witnessing increasing human populations demanding more energy and more food together with a more uncertain climate has led to a complicated dynamic interconnected web of physical, economic, and social components with many opportunities for intelligent adaptive management interventions. These interventions that change the distribution of water quantities and qualities over time and space can result in substantial economic, environmental, and social benefits. They can also introduce unexpected costs and risks. The constraints are physical (as with the large inputs of energy required for desalination), geographical (depending on the available suitable locations for reservoirs), financial (building, operating and maintaining infrastructure required to manage water is expensive), political (nobody wants to relinquish rights to scarce water without compensation), and ethical (what uses deserve to be prioritized, and how they relate to the needs of the environment).

Trade-offs are fundamental when allocating water to various sectors of society. Water is linked to the production of energy, food, industrial products and to human health and the condition of the broader environment. For many kinds of water uses, allocating water to one use usually means less water available for other uses. Consumptive use for agriculture, industry, or cities almost always involves trade-offs, as do mandates for instream flows to protect ecosystems or fisheries. But even consumptive uses do not diminish the total amount of global water. Consumption shifts water to a different part of the hydrological cycle: for example, from liquid to vapor, from clean to contaminated, or from fresh to salty.

Choices about managing water trade-offs involve more than hydrology and economics. They involve people's values, ethics, and priorities that have evolved and been embedded in societies over thousands of years. The juxtaposition of hydrology, economics, and values is at the crux of the water–climate–food–energy–environmental and society (people) nexus. While it is unreasonable to think that models of water resource systems will or even should include each component of this interconnected interdependent nexus of components, analysts must be cognizant that the part of system that they model is interacting with and being influenced by those components assumed exogenous to the system.

13.2   Water Resources System Components, Functions, and Decisions

13.2.1   Components

For the purposes of planning and management water resource systems include three components:

- The *natural resource system* (NRS) component consists of the streams, rivers, lakes, and their embankments and bottoms, and the groundwater aquifers, and the water itself. This includes the abiotic or physical, biological, and chemical ("ABC") components in and above the soil. It also includes the infrastructure needed to collect, store, treat, and transport water such as canals, reservoirs, dams, weirs, sluices, wells, pumping stations, pipes, sewers, and water and wastewater treatment plants, and the policies or rules for operating them.
- The *socioeconomic system* (SES) component is the water using and water-related human activities. This component can also include the stakeholders, i.e., the interested and affected public.
- The *administrative and institutional system* (AIS) component are the institutions that are responsible for the administration, legislation and regulation of the supply (NRS) and the demand (SES) components of the water resource system (WRS). This component includes those institutions that plan and build and operate the infrastructure required to insure that water is where and when and in the condition needed in ways beneficial to society.

13.2.2 Functions

Table 13.1 presents a framework of water resource system functions. This framework distinguishes between tangible and intangible functions. Tangible functions can be described quantitatively. For example, hydropower generation or municipal water supply, may be assigned a monetary value. Intangible functions are activities such as nature conservation or preserving a beautiful view that are hard to quantify in monetary terms. In between are environmental functions, some of which may be given quantitative values and others valued only indirectly, such as by using the opportunity cost associated with a particular target. The self-purification process of a river, for example, may be assigned a value by comparing this "work done by nature" with the costs of the least cost alternative that accomplishes the same results, such as constructing, maintaining and operating a wastewater collection and treatment system.
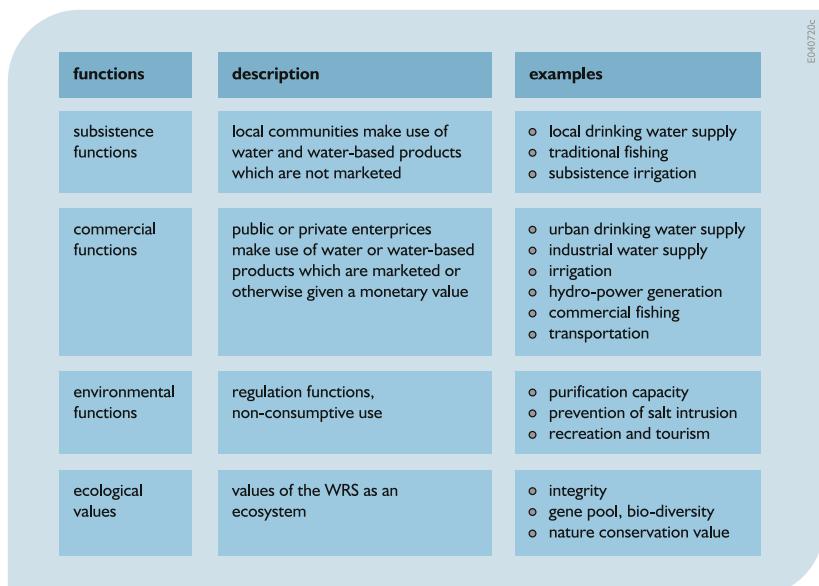
13.2.2.1 Subsistence Functions

Communities depend to a large extent on water for household uses, and for irrigating home gardens and community outdoor green and recreation areas. They may also use 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 or services, this may well be considered an economic loss.

13.2.2.2 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

Table 13.1 Functions of the water resources system

functions	description	examples
subsistence functions	local communities make use of water and water-based products which are not marketed	◦ local drinking water supply ◦ traditional fishing ◦ subsistence irrigation
commercial functions	public or private enterprices make use of water or water-based products which are marketed or otherwise given a monetary value	◦ urban drinking water supply ◦ industrial water supply ◦ irrigation ◦ hydro-power generation ◦ commercial fishing ◦ transportation
environmental functions	regulation functions, non-consumptive use	◦ purification capacity ◦ prevention of salt intrusion ◦ recreation and tourism
ecological values	values of the WRS as an ecosystem	◦ integrity ◦ gene pool, bio-diversity ◦ nature conservation value

"nonconsumptive use" should be regarded with certain reservations. Nonconsumptive water use may alter the performance of the WRS in various ways. For example, consider reservoirs built for hydropower. Reduced sediment and fish passage and increased evaporation losses may impact downstream ecosystems and users. Second, operation of the reservoirs for the production of "peak power" may alter the flow regimes downstream, and this can adversely affect downstream ecological habitats and users. Finally, water quality problems resulting from reservoirs may impact users and ecosystems. Another example of partly nonconsumptive 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.

13.2.2.3　Environmental Functions

The drainage basin of a river fulfills a series of environmental functions that require no human intervention, and thus have no need of regulatory systems. These functions include self-purification of the water and recreational and tourism uses. It is sometimes difficult to assign values to environmental functions. They may be assessed by using opportunity costs, calculated as the costs of providing similar functions in other ways, e.g., the cost of additional wastewater treatment. Lower bounds on recreational and tourism values may be estimated by assessing the economic benefits accruing from the use of tourist facilities including hotels, and/or the revenue obtained from the sale of fishing licenses.

13.2.2.4　Ecological Functions

Rivers, streams, and lakes and their associated wetlands, floodplains, and marshes offer an environment for aquatic species. Land–water ecotones (transition areas between adjacent ecological communities) are known to harbor a rich assemblage of species, and are also important for the diversity of adjacent ecological communities. These ecological entities have an intrinsic ecological value irrespective of actual or potential human use. There are many concepts and expressions that describe this ecological value: "heritage value," "aesthetic value," "nature value," "option value," "existence value," among others.

> **Box 13.1. Definitions**
> **Policy goal**: what do we want to accomplish?
>   **Strategy**: how do we want to do it?
>   **Decision**: what are we going to do?
>   **Scenario**: the external economic, environmental, or political situation affecting our strategy and decision.

13.2.3　Goals, Strategies, Decisions, and Scenarios

In planning projects the terms *goal*, *strategy*, *decision,* and *scenario* are frequently used. In popular use their distinction is often confusing. In this book we have used the following meanings:

- A *goal* defines what is to be achieved or how some target is to be met. Goals identify needs, prioritize issues and define targets and constraints on the actions to be taken to meet the targets. Goals may define preferred courses of action. For example, the goal might be to apply user-oriented demand management measures rather than relying on large-scale water supply infrastructure development.

- A *strategy* is defined as a logical combination of individual measures or decisions that accomplishes the stated goals and satisfies the constraints imposed on the WRS. For example, the construction of a reservoir plus the widening of the canal downstream and the increase of the intakes of the irrigation system all in an effort to reduce the risk of damage to the agriculture sector in a drought prone area is one strategy. An alternative strategy might be to implement a cropping pattern that uses less water.
- A *decision* is the implementation of a particular strategy or course of action. A distinction can be made between:
  - *Technical (structural) measures*: modifications of elements of the water resources infrastructure such as canals, pumping stations, reservoirs, and fish ladders. Technical measures often include managerial measures such as better ways of using the infrastructure.
  - *Ecological (nonstructural) measures* to improve the functioning of ecosystems, for example, by introducing fish fry in spawning areas, or large herbivores.
  - *Economical measures* to induce water consumers to alter their use of water by changing the price of the resource use (through charges, taxes, or subsidies).
  - *Regulatory measures* to alter the use of water (through land-use zoning, permits, pollution control and other forms of restrictive legislation).
  - *Institutional measures* specifying which governmental agencies are responsible for which functions of the WRS, and specifying the necessary interactions between the public and private sectors involved.
- A *scenario* is defined as the environment exogenous to the water system under consideration that cannot be controlled. Examples of scenario variables include rainfall and other aspects of the climate, demographical trends and changes, production functions (including crop water requirements), and most economic
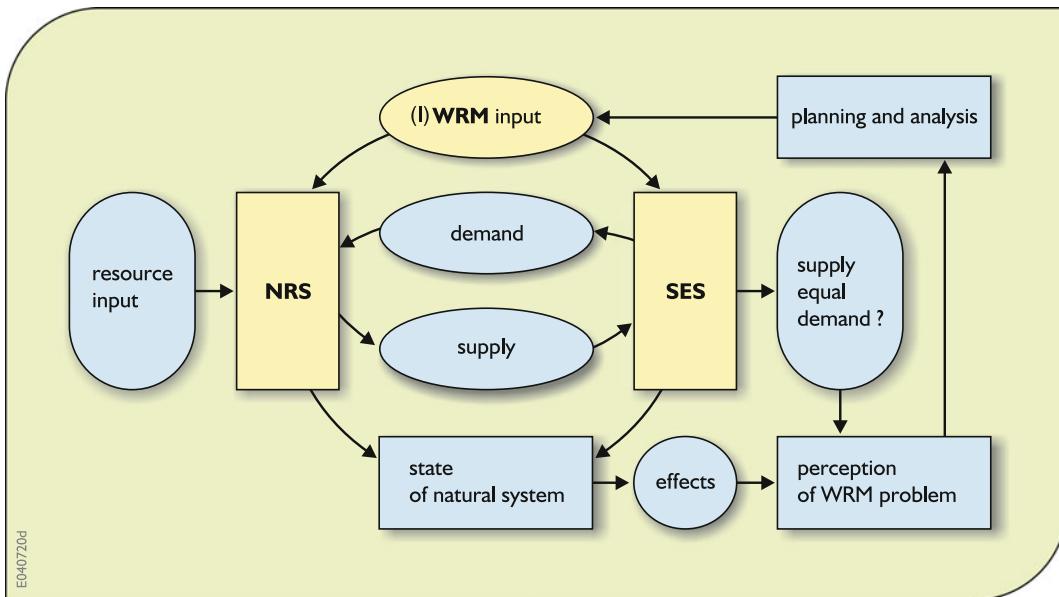
variables relating to benefits and costs. What should be considered as a scenario and what as a decision variable may depend on the system boundaries that have been defined.

13.2.4 Systems Approaches to WRS Planning and Decision Making

Literature on the systems approach to planning often emphasizes the mathematical techniques used by practitioners of this approach. This book is no exception. Most of it is devoted to modeling water systems. The use of mathematical tools, however, is only part of what constitutes a systems approach. The approach applied to complex systems of many interdependent components, involves:

- building predictive models to explain system behavior,
- 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 during these exchanges between analysts, planners, and decision-makers and stakeholders, and
- monitoring and evaluating the results of the strategies implemented.

Systems analysis and policy analysis are often considered as being the same. If a distinction is to be made, one might define systems analysis as being applicable to more than just policy issues or problems. It can be applied to any system one wants to analyze for whatever reason. System diagrams or conceptual models identifying system components and their linkages are important tools in systems analysis. A system diagram

Fig. 13.1 Identification of a water resources management (WRM) problem

represents cause–effect relations among the components of the overall system. An example of the use of system diagrams in analyzing water resources problems is presented in Fig. 13.1.

As Fig. 13.1 shows, water using activities may face two problems. First, the quantity demanded may be greater than the supply; second, they adversely impact the natural system (e.g., generate pollution or alter the water level). The perception of these problems can motivate analysis and planning activities, which in turn can result in management actions. The figure shows that the problems can be addressed in two ways: either by implementing demand-oriented measures (addressing the water use, i.e., SES), or by developing infrastructure that impacts the NRS). Demand-oriented measures aim to reduce water use and effluent discharge per unit of output. Supply-oriented measures on the other hand are aimed at increasing the water supply so that the magnitude and frequency of 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 selected by the implementing authority.
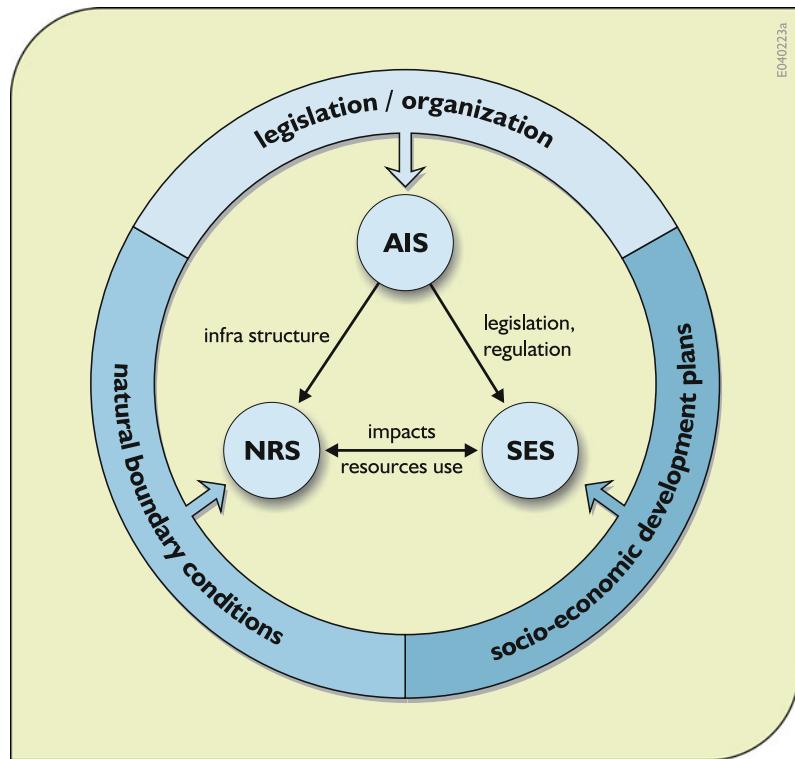
13.3 Conceptual Description of WRS

Water resources management aims to increase the benefits to society from the existence and use of water (NRS). Just how best to do it is society's (SES) choice, commonly made through its governing institutions (AIS). These three "entities" are depicted in Fig. 13.2.

The management actions among the components of a WRS system are depicted by the arrows shown in Fig. 13.2. The arrows represent only the actions, not the information flows. There must be information feedbacks, otherwise effective management would be impossible. Each of the three systems is embedded within its own environment. The NRS is bounded by climate and 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 operates within. Boundary conditions are usually considered fixed, but in some cases they may not be. For example, climatic conditions

Fig. 13.2 Context for water resources planning involving the natural resource, socioeconomic and administrative–institutional systems

may be considered to be changing due to global warming. Similarly for laws and regulations. Whether and, if so, when to consider the possibility of changes in this "external" environment should be decided at the start of any planning project.

Consider, for example, regional economic . This predicted growth is often treated as given. If the water resources available cannot sustain this projected growth (or only at very high costs), it may be appropriate to reconsider this assumed growth. By learning the consequences of unrestricted growth at the regional level, planners can consider the desirability of other options that might be considered at higher (usually national) planning levels. This is represented in Fig. 13.2 by the border frame "socioeconomic development plans". In fact, the arrow pointing inwards to the SES is reversed in such a case: the analysis provides information to a higher planning level that can change the boundary conditions.

13.3.1 Characteristics of the Natural Resources System

The natural resources system (NRS) is defined by its boundaries, its processes, and its control measures.

13.3.1.1 System Boundaries
The study area of a planning project will often coincide with an administrative boundary (state, county, district, province, etc.). However, a WRS is typically defined by its hydrologic boundary. These political and hydrologic boundaries can differ. Clearly, any planning project for a WRS must include the larger of these boundaries, but not necessarily everything within them depending on the purpose of the study and the particular WRS. The consideration of problem sheds that contain the components that impact water sheds is often useful.

For the purposes of modeling it has often proven 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 area boundaries can be reasonably determined and modeled.

13.3.1.2 Physical, Chemical, and Biological Characteristics

The *physical processes* in an NRS are transport and storage within and between its subsystems. For the surface water system, a distinction is usually 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 *biological* and *chemical characteristics* define the biological and chemical composition of groundwater and surface water and the transport, degradation and adsorption processes that may influence this composition. The level of detail to which these characteristics are considered will depend on the requirements and threats they impose on the water using and water-based activities.

13.3.1.3 Control Measures

By adding or changing the values of system parameters defining design and operating policy options of NRS, water resources managers can change the state of the system. An example is the rule curve defining how much water to release and when for different purposes. Another example is the flow capacity of feeder canals. Increasing the capacity of these canals permits greater allocations of water to farmers. An example of nonphysical 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.

13.3.2 Characteristics of the Socioeconomic System

Like the NRS, the SES has its boundaries, processes, and control measures.

13.3.2.1 System Boundaries

The economic and social system generally does not have a physical boundary like that of the natural system. Economic and social activities in a river basin, for example, are connected to the world outside that basin through the exchange of goods, people, and services. The factors that determine the socioeconomic activities to include in a project planning exercise will depend on the context of the problems and development opportunities being considered. Outside the boundary of the socioeconomic system are factors or conditions that are beyond the control of the WRS decision-makers.

13.3.2.2 System Elements and Parameters

The socioeconomic part of the WRS can be defined by identifying the main water using and water-related activities, the expected changes and developments in the study area, and the parameters whose values define these changes and 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 SES include:

- 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 demands for electric energy.
- Public water supply: location of centers of population and industrial activities, expected growth, alternative resources.
- Recreation: nature and location, expected and desired development, water quality conditions.

- Navigation: water depths in relevant parts of the open channel system.
- Nature conservation: location of valuable and vulnerable areas and their dependence on water quality and quantity regimes.

Some examples of important system parameters of the SES are labor 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 analyzing activities in the study area, it is important to consider possible discrepancies between the opinions of individual actors or stakeholders and their representatives. For example, individual farmers may have different interests than suggested by the official agricultural organizations.

13.3.2.3   Control 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. In such cases, the price of water (or taxation for waste water discharges) may not be an acceptable control variable (except perhaps to inform stakeholders on the consequences of possible cost reduction measures).

13.3.3   Characteristics of the Administrative and Institutional System

The AIS, like the NRS and SES, has its boundaries (its authority or limits) and its processes including its ways of reorganizing for improved performance.

13.3.3.1   System Elements

Administrative and institutional settings vary with scale, and with the way governing institutions exist and operate. In many countries, but certainly not all, the institutional framework consists of:

- 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 districts.

When initiating broad comprehensive water planning projects knowing the following information is useful:

- the ministries and coordinating bodies having authority and responsibilities related to water resources management
- the agencies involved in the preparation of water resources development plans
- existing national and regional water resources development plans and the authorities responsible for implementing these plans, establishing and enforcing regulations, and overseeing infrastructure construction and operation
- 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 often useful information includes the policies and plans of various water-related sectors such as environment, agriculture, economy, transportation, urban development and energy.

13.3.3.2   Control Measures

From a systems point of view, the decision or control variables that can be changed in the AIS are less clear than in the case of the NRS and SES. Often measures can be taken to improve the functioning of the system, for example, by establishing coordinating bodies when these are not present, shifting responsibilities toward lower levels of government, privatization, and other measures. If they cannot be changed, at least possible beneficial changes can be identified and presented to those responsible for making decisions.

13.4   Framework for Analysis and Implementation

A water resources planning study generally comprises five general phases, as illustrated in Fig. 13.3. Although we do not suggest the use of any 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 among the activities in them, is referred to as the analysis framework. A coherent set of models is typically used for the quantitative analyses aimed at identifying and evaluating alternative beneficial measures and strategies.

A decision process is not a simple linear sequence of steps as suggested in Fig. 13.3, but involves feedbacks to earlier steps. Part of the process is thus iterative. Feedback loops are needed when:

- solutions fail to meet current criteria
- new insights change the perception of the problem and its solutions
- essential system components and links have been overlooked
- goals and objectives or the scope of study change (e.g., due to changing political, international, developments in society).

Communication and interaction with the decision-makers are essential throughout the duration of a planning project and the implementation of the selected development. To ignore this increases the risk of generating plans and policies that are no longer relevant or of interest to the client. Regular reporting (inception and interim reports, etc.) helps in effective communication, but a continuous dialogue is important throughout all stages or phases of the analysis.
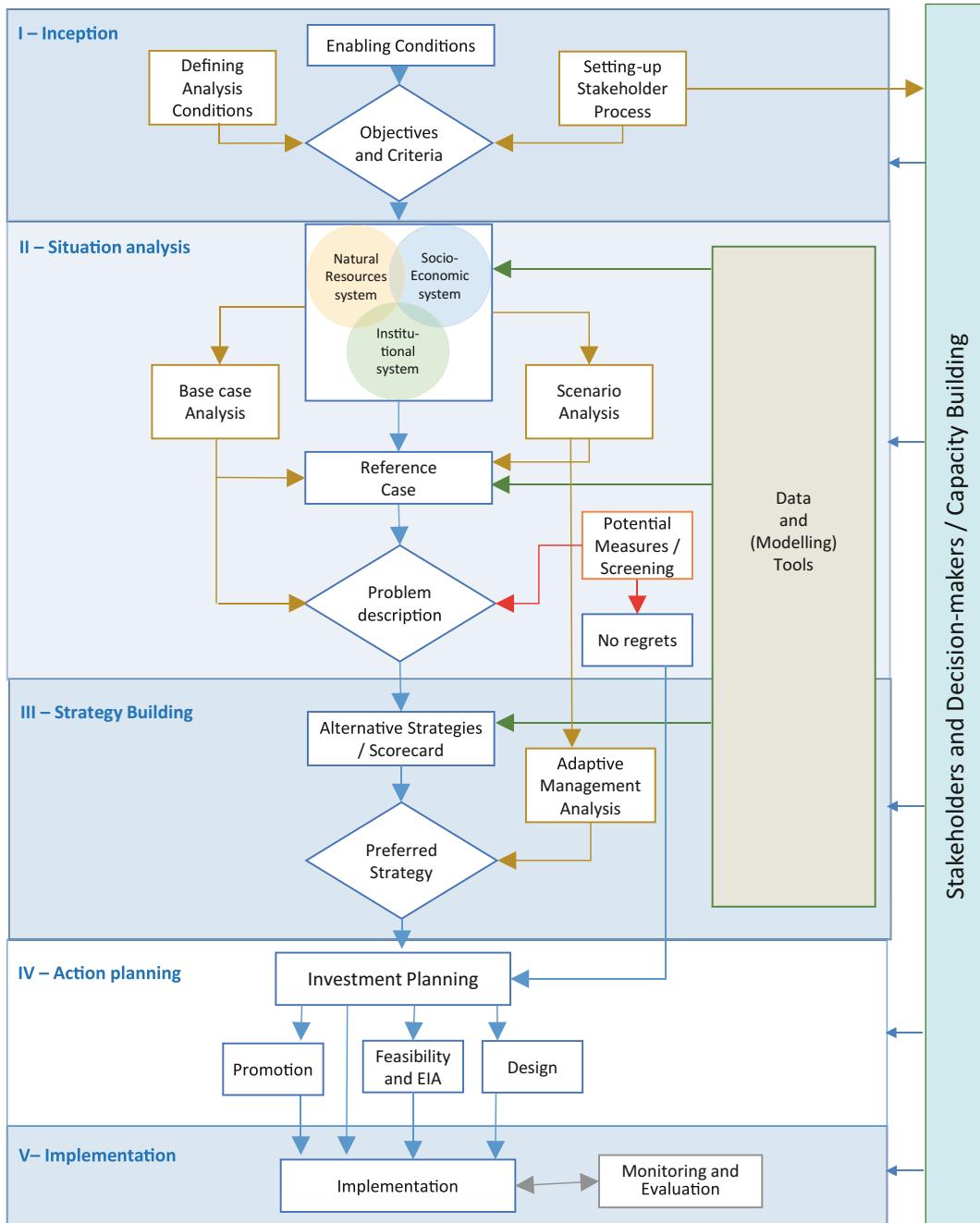
Decision makers and stakeholders should be involved in each of the five (idealized) stages of this framework. Otherwise there is a risk of the planning project producing results that those potentially impacted will not support. Stakeholder involvement brings both knowledge and preferences to the planning process—a process that typically will need to find suitable compromises among all decision-makers and stakeholders if a consensus is to be reached.

The framework involves a series of decisions at the end of each stage. The divergence–convergence process for involving stakeholders in decision-making on the five analysis stages is illustrated in the rhombus approach of Fig. 13.4.

The first *inception stage* of the process identifies the subject of the analysis (what is to be analyzed and under what conditions), the objectives (the desired results of the analysis), and constraints (its limitations). On the basis of this analysis, during which intensive communication with the decision-makers is essential, an agreement on the approach for the remainder of the analysis needs to be achieved. The results of the inception stage can be presented in an inception report, which includes the work plan for the other phases of the analysis.
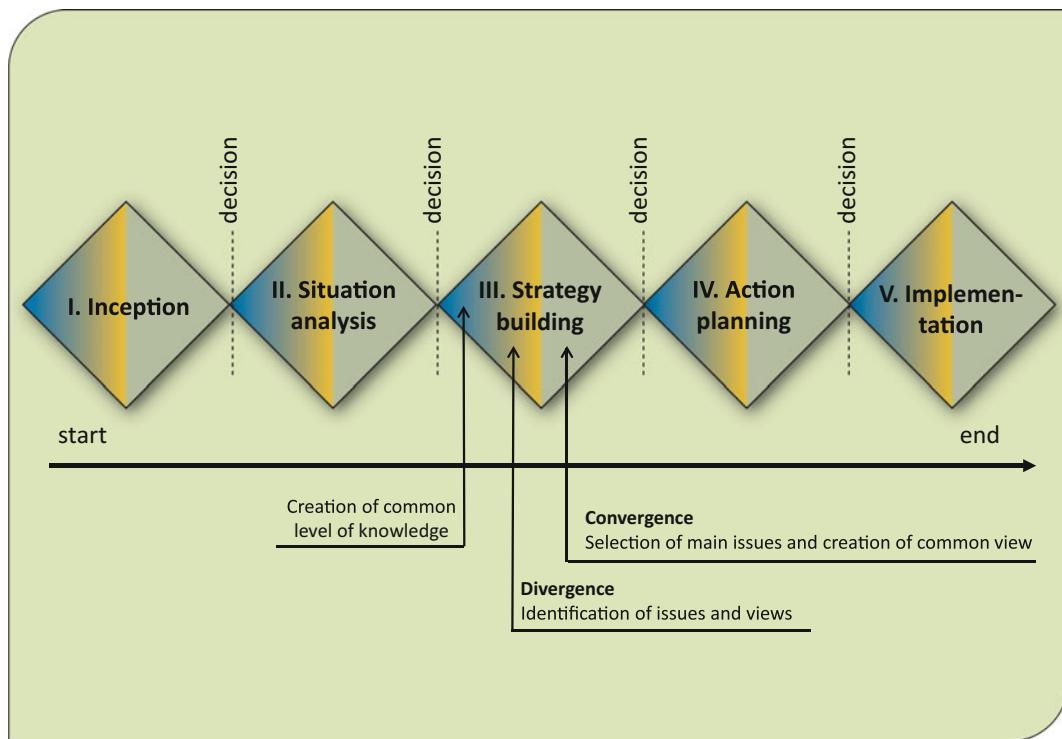
In the *situational analysis stage*, the tools for the analysis of the water resource system are selected or developed. Major activities in this phase typically include data collection and modeling. The models will be used to quantify the present and future problems in the system. Scenarios will be developed to describe the future boundary conditions for the system. Identifying and screening of alternative decisions can occur in this phase. If possible no regret measures will be identified for immediate implementation. A gradual improvement of the

Fig. 13.3 Framework for analysis and implementation of water resources projects

understanding of various characteristics of the WRS is often obtained as the study progresses from limited data sets and simple tools to more detailed data and models. Interaction with the decision-makers will be greatly enhanced if they or those they trust and communicate with are involved as part of the analysis team. More formal interaction can be structured through presentations of results in meetings and in interim progress reports.

Fig. 13.4 Divergence—convergence process in decision-making

In the *strategy building stage* alternative strategies will be developed and discussed with the decision makers/stakeholders. This will include adaptive management elements to ensure that the preferred strategy is sufficient robust and flexible in case the future develops differently than expected.

In the *action planning stage* the selected strategy will be prepared for implementation. An implementation plan will be developed that describes what will be done, by who, how it will be financed, etc. This stage often requires also additional work on components of the strategy (such as feasibility and design studies), and environmental impact assessments (EIA). Promotion of the he selected strategy is needed to "sell" the proposed measures to public. Finally, institutional arrangements will have to be made to ensure a smooth implementation.

Finally, in the *implementation stage* the actual implementation will take place. Continuous monitoring and evaluation is needed to adjust the implementation plan when this appears to be needed, for example, because the conditions (e.g., finances, social pressures, political mood) change.

Each stage or phase needs to provide the information desired by those institutions who will decide on what is best to do, and when, and how. What those governing institutions need to know to be better informed before making their decisions will of course vary among different planning projects. But whatever that information is the purpose of performing analyses is to create and communicate it. The results of the analyses performed in a planning project should be of no surprise to those reading them in a final project report. Again, communication between the project and the requesting institutions, and the affected public—the stakeholders—is essential throughout the project. This communication may not guarantee a consensus but it can certainly help the project team in their efforts to find it.

13.4.1 Step I—Inception Phase

Water resources planning studies are often triggered by specific management problems such as the need to increase power production or water supply reliability, 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, data, and time). The initiators of the study generally have more or less concrete ideas about the objectives and purpose of the analysis. However, these can change during a study.

The client's ideas about the problems and issues to be addressed will usually be described in a Project Formulation Document (PFD) or Terms of Reference (ToR). The very first activity of the project is to review and discuss the contents of these documents. If the subject (what needs analyzing) and objectives (what is to be accomplished) are adequately described in the ToR, the next step of the study is to specify and agree on the approach (how).
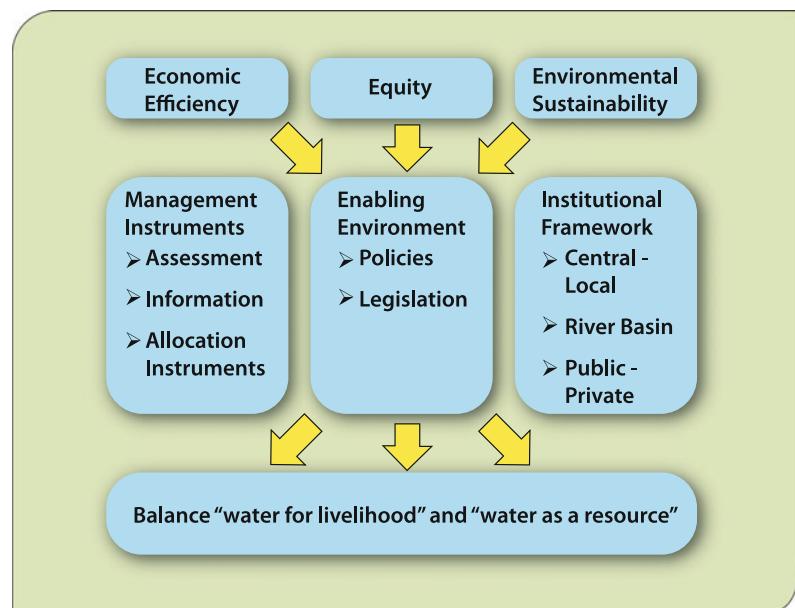
In many situations, however, the next task of the project will be to assist the decision-makers in further specifying the objectives and subject of the analysis. For this activity, intensive communication is required with authorities involved in water resources planning and the stakeholders. They can 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 (including just stalling for time hoping stakeholders will lose interest in a particular issue). Furthermore, objectives can change over time. As emphasized above, constant and effective communication between analysts and their clients is absolutely essential to the success of any planning project. We mention this often as it is not always easy given busy time schedules and often having to learn the differences in the meanings of various words or expressions (jargon) used by all parties.

13.4.1.1 The Enabling Conditions

In order to successfully carry out a good planning study certain conditions should be met. Most of these conditions are external to the project activities. This means that they should have been set before the planning exercise starts. A generic description of the enabling conditions for integrated planning is given in Background Paper no. 4 (GWP 2000) and is illustrated in Fig. 13.5.

Fig. 13.5 Enabling conditions (the "pillars") for IWRM

- Enabling environment at national level:
  - national water legislation and national policies that guide the planning process and enables enforcement.
- Institutional framework:
  - existence of water institutions at national and regional level with qualified staff;
  - in case of river basin studies, existence of some kind of river basin organization (RBO) at river basin level.
- Management instruments:
  - availability of data, information, and tools that enables informed decision making.

In the Inception stage it should be determined which conditions are relevant for the specific planning exercise. This depends on the issues involved. If needed, institutional measures can be part of the planning project.

13.4.1.2   Setting Up the Stakeholder Involvement Process

The very first step is to set up the stakeholder involvement process. Which stakeholders to involve and how will depend on the specific basin and the issues to be addressed. In general two categories of stakeholders can be identified:

- the people and organizations that will be affected by the plan; and
- the people and organization that are needed to implement the plan.

In some cases a stakeholder analysis might be needed to determine the best stakeholder involvement process. More detail on involving stakeholders is given in Sect. 13.5.1.

13.4.1.3   Defining Analysis Conditions

In addition to the more legal and institutional oriented conditions as described in Sect. 13.4.1.1 it is necessary to get agreement on the analysis conditions for the planning study. This includes:

- The base year for the study:
  - the most recent year for which basic data on the present situation is available;

- The time horizon(s) for the study:
  - this may include short term (e.g., 5 years), medium term (e.g., 20 years) and long term (>25 years);
- The discount rate to be applied in the economic analysis:
  - taken as specified by (e.g.) the Ministry of Finance or Economic Affairs, or by the financier of the planned investments (e.g., ADB, World Bank and JICA);
- System boundaries of NRS, SES, and AIS— the components and the level of detail that will be included:
  - e.g., will the coastal zone be included in a river basin study?
  - are the results to be presented at local government unit level?
- Time periods based on within- and over-year variability of systems processes and inputs
- Scenario assumptions concerning factors external to the WRS, such as the growth of population, food and energy consumption and prices. See also Sect. 13.4.2.4.
- 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.
- Data, time, and budget constraints. Studies have to be executed within constraints of available data, time, and budget.

The choice of the time horizon is often given insufficient attention. Official planning horizons (e.g., 5, 10, and 25 years) are typically used as time horizons for elements of the analysis. However, one should also consider the timescales of the system and the processes within it. System components will have characteristic time scales. For example:

- Economic activities have life cycles that are usually determined by the amortization period of the investments. Time horizons of planning processes can be 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 generally adopted to properly consider long-term effects in the evaluation process. Decision-makers tend to focus on short-range decisions even if they impose possible risks in the long term, because their political time horizons are often limited to (or renewable in) short terms and hence they prefer short-term political gains.

13.4.1.4 Objectives and Criteria

An essential activity in the inception phase is the translation of general objectives, as described in the ToR or in policy documents, into operational objectives that can be quantified. Examples of objectives and criteria are discussed throughout this book and especially in Chap. 9. 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 chapter.

National and regional development objectives
An essential component of an integrated plan is the connection of the plan and its objective to national development goals as well as to common international goals (e.g., the Sustainable Development Goals—SDGs). The plan should refer to national policy priorities and indicate the contribution the plan will make to the various development goals. Required information is usually described in various national policy documents. In addition to the national policy documents any existing regional/provincial policy documents need to be taken into account. Each plan need to have an agreed objective that not only focuses on the main, but also expresses the relation with above mentioned national and other sector plans, as well as the contribution the basin can make in realizing these higher level plans.

Operational objectives, criteria and targets
If needed, the general objectives as stated in the national policy documents have to be translated into operational objectives for the specific area under consideration, e.g., a river basin. This should be done by specifying them in socioeconomic terms, amongst others, which are meaningful to the decision makers and stakeholders. For each objective evaluation criteria should be defined as a measure of how far the defined objectives have been achieved and, if possible, clear targets should be specified. Monitoring will indicate how far the objectives have actually been achieved. This process, illustrated in Fig. 13.6, is discussed in more detail in Chap. 9.
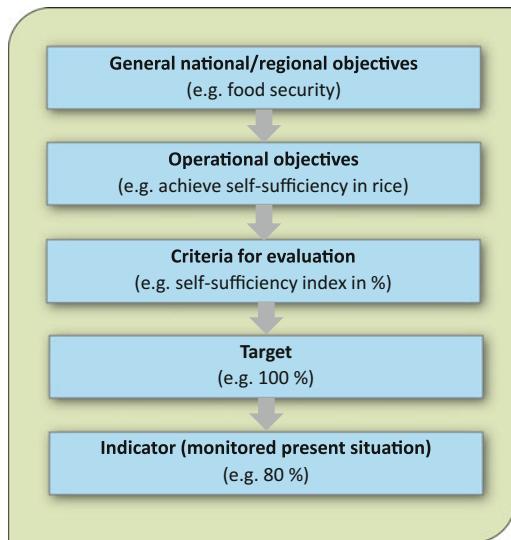
The evaluation criteria need to be comprehensive (i.e., sufficiently indicative of the degree to which the objective is achieved) and measurable. The criteria do not all have to be expressed in a single measurement scale. Criteria can be expressed in monetary and nonmonetary terms.

It may be useful to incorporate sustainability as an objective, and if so, it may also be useful to relate them to the UN Sustainable Development Goals (SDGs), the SDG targets and the indicators, that have been selected to monitor the SDGs.

Illustrative river basin case
Table 13.2 presents a scorecard that summarizes results of an analysis for a river basin case. The results of the Inception step (i.e., the objectives and criteria), for this river basin are given in the

Fig. 13.6 Making objectives operational

- Objective 1: Provide safe water and sanitation for the people;
  - % people access to safe drinking water;
  - % people access to sanitation facilities;
- Objective 2: Increase food production;
  - Irrigation area (ha);
  - Number of animal water points (#);
- Objective 3: Support economic sectors—industry and energy;
  - Water supplied to mining (% of demand);
  - Water supplied to industry (% of demand);
  - Hydropower generated (MWh);
- Objective 4: Protect the Environment;
  - Protected watershed area (km2);
  - Number of springs/sources protected (#);
  - Average class water quality rivers (class A to D);
- Objective 5: Decrease vulnerability to floods and droughts;
  - Vulnerability to floods—average damage ($/year);
  - Vulnerability to droughts—average damage ($/year).
- In addition two implementation-related criteria were formulated to evaluate the strategies:

first two columns of the table. They show that for this case five objectives were formulated. For each objective 2 or 3 criteria were identified that expresses in how far the objective is or will be achieved:

Table 13.2 Example of a scorecard showing objective values associated with various strategies

Objectives and criteria	unit	Base Year 2010	Targets 2020	Targets 2030	Perfect	Ref. case (no action) 2020	Ref. case (no action) 2030	Strategy 1 2020	Strategy 1 2030	Strategy 2 2020	Strategy 2 2030
Obj.1: Water and Sanitation											
% people access to safe drinking water	%	50%	63%	73%	100%	63%	73%	63%	73%	63%	73%
% people access to sanitation facilities	%	30%	50%	70%	100%	50%	70%	50%	70%	50%	70%
Obj.2: Food production											
Irrigation area	1000 ha	24	30	35	40	26	28	28	31	30	35
# animal water points	#	300	500	900	1000	400	700	500	900	500	900
Obj.3: Industry and Energy											
Water supplied to mining	%	30%	80%	90%	100%	40%	50%	50%	70%	80%	90%
Water supplied to industry	%	70%	80%	90%	100%	70%	70%	80%	90%	80%	90%
Hydropower generated	MWh	34	80	120	120	34	34	70	110	80	120
Obj.4: Environment											
Protected watershed area	km^2	1200	2500	3500	3500	2000	2500	2500	3000	2500	3500
Number of springs/sources protected	#	300	600	900	900	400	600	500	700	600	850
Average class water quality rivers	I - V	II	III	IV	V	II	III	III	III	III	IV
Obj.5: Vulnerability											
Vulnerability to floods - average damage	m€/yr	120	< 78	< 50	0	100	80	100	80	78	50
Vulnerability to droughts - average damage	m€/yr	200	< 50	< 30	0	160	120	80	40	50	30
Implementation information											
Required investments	m€		-	-	-			400	650	600	1200
B/C ratio economic categories (Obj.2, Obj.3)	-		> 1,3	> 1,2	-			1.3	1.2	1.2	1.1

– Required investments ($);
– Benefit/Cost ratios of economic categories (</>).

13.4.1.5   Work Plan and Decision-Making

Once it is clear "what" will, as well as what will not, be analyzed and "why", analysts can specify "how" this will be done. A description of the system to be analyzed includes the conditions and the assumptions under which the analysis will be performed.

All required activities can be combined in a work plan. It is often advantageous to develop a critical path network of the various analysis tasks. Critical path networks define the sequence of various tasks required to complete an analysis, or indeed the entire planning project, and their start and finish times. This will guide the allocation of personnel and identify the time needed to perform such tasks. These networks can be updated as the project proceeds. Such networks are useful for scheduling activities and personnel involved in the project, and for ensuring (or at least increasing the probability) that data and personnel will be available for each activity when needed and when decision-makers and stakeholders are to be involved in the analyses or in workshops or meetings focused on improved understanding of project progress and goals.

Data Availability

An important boundary condition for studies is often the availability of data and other information required for the study. 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 depend on the problems to be addressed and the objectives to be satisfied.

Level of detail One of the main tasks of a project leader is to motivate and manage the experts from various disciplines. 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 will have to obtain the additional detail needed fulfill the objectives of the analyses. Sometimes the right level of detail is chosen, but team members may be tempted to spend too much time addressing more detailed questions of interest to them and fail to come up with the information desired within the available time. Maintaining the proper level of detail is one of the main reasons for feedback loops in the analysis process.

Computational Requirements

An important element of the work plan will be the determination of the computational resources needed for the analysis. This includes mathematical models, databases, GIS, and the like. Together these must be used in a way that describes the system and permits an evaluation of possible measures and strategies under different scenarios at the level of detail desired. Often a combination of simulation and optimization models has proven useful.

For the purposes of analysis, the study area is typically subdivided over space and time into smaller units considered to be homogeneous with respect to their characteristic parameters. Each unit can be included in mathematical model(s). The number of elements required for the analysis depends on the issues being addressed, the complexity of the study area, the 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 or temporal detail is required model elements can be subdivided. The assumptions and conditions under which analyses are undertaken should be specified in close cooperation with those institutions overseeing and contributing to the study.

Work Plan

The results of the inception phase are documented in an inception report. This report can serve as a reference during the execution of the study. An essential part of the report is the proposed work plan, in which time, budget and human resource allocations to various activities

are specified. This work plan typically includes bar charts (possibly derived from critical path analyses) for activities and staffing, time schedules for deliverables, milestones, reporting procedures and similar features. The report should include a communication plan that describes the interaction between the decision-makers and stakeholders and the analysis team.

Inception Report

An inception report is a specific and concrete result of the inception phase. It contains the findings of and decisions made during the inception phase. It should make clear what will be studied, and 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. It is an important product because it contains all that has been learned in this first inception 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 between the analyst and the client that took place during this phase. It should state the client's views about problems, objectives and other aspects. Project analysts must understand the client's concerns, problems and objectives. 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 planners, analysts or consultants. To achieve such ownership, frequent interaction must have taken place among the analysts, the decision-makers and stakeholders, to a much greater extent than is indicated in Fig. 13.3. This can be done in specific workshops, such as those devoted to the problem statement or to the specification of objectives and criteria.

13.4.2   Step II—Situation Analysis

In the situation analysis phase the study starts to dig deeper in the water resource system. Its various components will be studied in detail, data

will be collected and where necessary and possible the system components will be captioned in models. As much as possible this should be done in close collaboration with the stakeholders to ensure that the analysts and stakeholders have the same understanding of the system. Once these models are available a structured analysis can be carried out to quantify the present and future problems and a start can be made with identifying measures to address these problems.
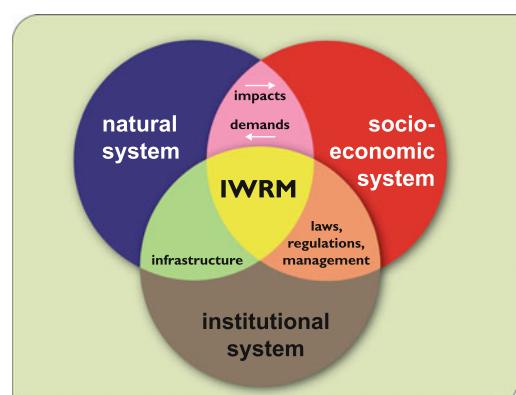
13.4.2.1   Understanding and Describing the Water Resources System

A WRS comprises:

- Natural (Resources) System (NRS);
- Socioeconomic System (SES); and
- Administrative and Institutional System (AIS).

Each of the three systems is embedded within its own environment. The Natural Resources System is bounded by climate and (geo)physical conditions. The SES is formed by the demographic, social and economic conditions of the surrounding economies. The AIS is formed and bounded by the constitutional, legal and political system. The interlinkages of the three systems are illustrated in Fig. 13.7.

It is important that the plan includes a good description of the integrated elements of the

Fig. 13.7 Systems components of a WRS

WRS. Most decision-makers and stakeholders will be nontechnical or only know about a limited part of the overall system. To be able to make balanced decisions they should understand how the overall system functions and how interventions in one part of the system will impact other systems elements.

The situational analysis starts with an inventory of the *characteristics* of the WRS. This requires the reduction of a complex reality into a comprehensible description of system components and linkages. Choices have to be made about what (the detail that) should be included and what can be ignored. Such choices require engineering and economic judgment in combination with an understanding of the problems and possible measures that can be taken to improve system performance. The next step will be an inventory of the *activities* and *ongoing developments* that will determine how the system will perform in the future and what kind of additional activities can be expected. This can include autonomous developments (such as population and urban growth) as well as policy decisions that have been or may be taken that could influence the characteristics and performance of the WRS. An inventory of *policies and institutions* is helpful for identifying who is involved in the management and development of the system (and hence who should be involved in the analyses) and their objectives and opinions. This knowledge will contribute to the development of scenarios for the analyses.

Analysis of the Natural Resources System (NRS)

The NRS comprises the natural and engineered 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 water withdrawals to satisfy the demands of water-using activities. Such models have been introduced in many of the previous chapters of this book.

The results of the water quantity modeling may be the inputs for water quality models. 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 should have been selected in the inception phase. The analysis of the biological system aims to determine the response of the ecosystems to water resources management (see Chap. 10). Since often there is too little exact information on individual biotic components and their behavior under different hydrologic and chemical regimes, models of ecosystems typically depend on habitat parameters.

Analysis of the Socioeconomic System (SES)

Developments in the SES determine the way demands on the NRS may change. 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 activities adjacent to a reservoir. These SES developments in turn increase the water demands. Economists or planners may be able to estimate future levels of the activities dependent on water discharges and storage levels. These relations can be incorporated into water resource planning models.

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 can help provide information on factors that have been decisive in bringing about the present situation and that may give clues about the likely impacts of future developments. One's attention should be on the most important factors that determine relevant water-related activities rather than on analyses of the total economy. However, the difficulty in forecasting economic development is the uncertainty about which factors will be decisive for this development.

Part of the data needed to develop planning models is the relation between the economic

activities and their water use. Data are needed that define the type and amount of water used by various activities. Data are needed identify the following with respect to each identified activity:

- the amounts of water (quantity and quality) demanded and consumed during which periods of the year and at which locations
- the amounts of water discharged and the pollution loads during which periods of the year and at which locations
- the benefits to the user if these amounts are made available
- the damage to the user if these amounts are not available
- costs that can be recovered by having the user pay for the water and its influence (both at the intake and the discharge sites of his activity) on the water use pattern.

All these data should be able to contribute to the estimates of future water demands, consumption and wastewater discharges per unit of activity. As well as the level of activities and the resulting water demands, knowledge of the geographical location of water using activities (the pattern of activities) is necessary. 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 and their water use characteristics are unknown, it may be necessary to study the water use characteristics of similar activities in other regions.

The resulting water demand data need not always 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 person or unit of product. If supplies and demands are matched before the effects of such incentives are analyzed then one may over estimate needed capacities, because the "given" demands may be lower if water users are confronted with the costs as well as the benefits of water use. This type of internal feedback should be considered in the study.

Future water demands are often dependent on future scenarios. A water demand scenario is a logical but assumed combination of basic SES parameters and their effects on water-related activities, including the resulting water demands. An understanding of the functioning of the SES developed through the assessment of past and present trends is often helpful when formulating a limited number of consistent scenarios. Box 13.2 is an example of one such scenario.

Box 13.2. Example demand scenario

The water demand in an agricultural area depends largely on the availability of land and the crops being irrigated. 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 the agriculture sector 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 political pressure for agricultural development to meet this objective may be considerable. The water demand corresponding to this desired agricultural development could show the need for further development of the water resources in the region.

Analysis of the Administrative and Institutional System (AIS)

An analysis of the AIS is required to identify any legal or regulatory or institutional constraints on water resources management. Attention must be given to the interaction between various authorities involved in water resources management and to the effectiveness of the AIS. Arrangements made in the past concerning the use of water (water rights) should be identified, 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 the agency can implement effectively. The responsible agency should be aware of the possible role they may have in solving the management problems. Sometimes, the analysis of the AIS may result in recommendations for institutional and legal changes.

13.4.2.2   Data and Modeling

The result of the data collection and modeling activities is a quantitative representation of the WRS at an appropriate level of detail. The framework is designed to assess the effects of individual measures or combinations of measures, expressed in values for the evaluation criteria chosen. If computer programs for running models have to be developed or if existing computer programs have to be adapted in a significant way, a considerable effort may be required which 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 project, is important.

During the modeling activity, more information on the study area and the type of measures to be considered may become available. This could lead to changes in model structure. The models should therefore be flexible and adaptable to new information.

Model Integration

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. For example, the output of a water quantity model may be the input to a water quality model requiring different spatial and temporal resolutions. Some models may include links to various sub-models and run interactively, others not. Depending on the models and the problem situation, single or multiple linked models may be included within an interactive decision support system. In other cases, a clear description of information flow from one independent model to another may be sufficient.

Figure 13.8 provides an example in which various simulation models are combined to analyze a river basin under drought conditions. The reservoirs in the system involve sedimentation and hydropower generation. The core of this modeling framework is formed by the "core models" block in the upper right corner of the figure. In this block the demand for water is determined, followed by a balancing of supply through water allocation decisions. Links among these core models are automatic. Other models are linked through file transfer. This applies to the required input on macroeconomic and hydrometeorological conditions (generated by scenarios) as well as the side analysis 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 a multi-criteria analysis.
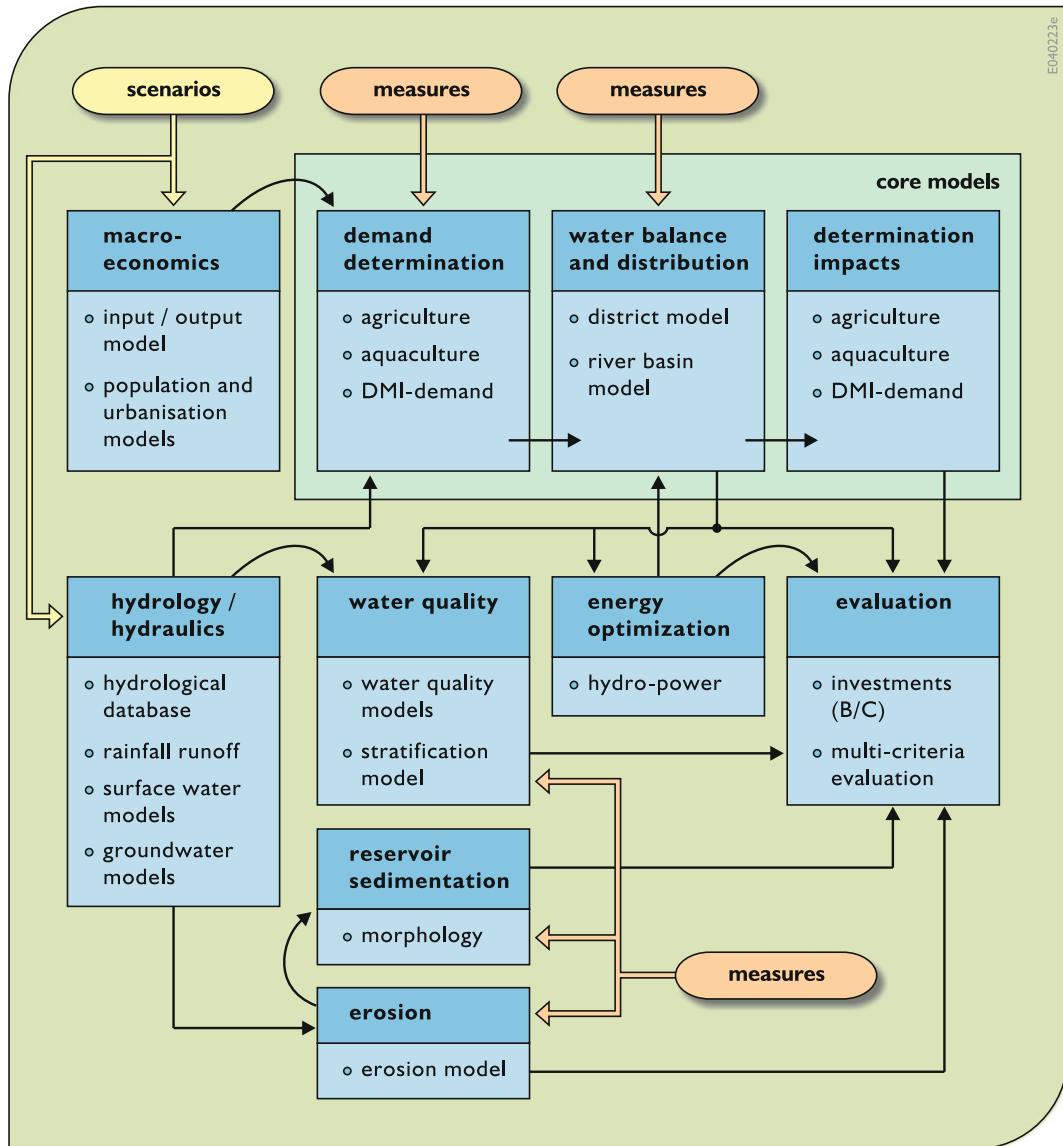
At various places in this modeling framework, one can change the values of input parameters. Scenarios can be analyzed by changing the macroeconomic and hydrometeorological conditions.

Figure 13.8 is just an example. Other problem situations may require different modeling frameworks. The goal in creating such model frameworks is to make them as simple and transparent as possible, and still adequately address the problems to be solved. Sometimes complexity is necessary. In any event it saves time and money to start as simple as possible and only add more detail when necessary to carry out a proper analysis.

Collaborative modeling

Involving decision-makers and stakeholders in the analysis process has till recently been limited to the more general analysis about problems and solutions. The quantitative information, e.g. resulting from models, was provided by the analysts (e.g., consultants) as input for the discussions. More and more we see that stakeholders do not accept this black box approach anymore. They want to understand what went into the model, how the models work and, preferably, they want to "play" with the model themselves. This is a promising development as this will increase the

Fig. 13.8 Example of typical computational framework of simulation models

understanding of the stakeholders on how the system works and let them see the opportunities and constraints of that system. Having stakeholders involved in the development and running of the models requires that these models are made more accessible and intuitive, in particular their input/output interfaces. It requires also a different attitude of the modelers. Various approaches to collaborative modeling are currently being developed, sometimes under different names such as Collaborative Modeling for Decision Support (e.g., shared vision modeling), Mediated Modeling, Group Model Building, Companion Modeling, Interactive Modeling, Networked Environments for Stakeholder Participation or Model-supported Collaborative Planning.

13.4.2.3 The Need for a Structured Quantified Analysis Process

Decision making on measures and strategies to improve the performance of the WRS should be based on quantified information about the present problems (e.g., average flood damage) and the impacts of proposed measures (e.g., the reduction in flood damage) and the costs of these measures. To be able to produce this quantified information the following is needed:

- a structured analysis process (this section); and
- a computational framework (see previous section).

The analysis process starts with a quantified problem description. The analysis of the present situation is called the Base Case analysis. To be able to predict possible future problems scenarios should be defined on how this future might develop. The computational framework will calculate the impacts (the future problems) of these possible external developments. This is often called the Reference Case analysis.

Base case

The performance of the WRS is studied for the infrastructure and water demands in the base case. The base case is based on the base year, which is the most recent year for which a complete set of data can be collected. The base case describes thus the performance of the WRS in the present situation. A comparison of the base case with the criteria (and possible targets) specified in the WRM objectives will result in a quantified problem statement.

Scenario conditions

A good plan should also address the expected water-related problems in the future. The analysis for the future time horizon(s) should include different scenario conditions. Possible scenario conditions for WRM are socioeconomic developments (change in demand and pollution) and climate change (including sea level rise). See the next section on more information about developing scenarios.

Reference case
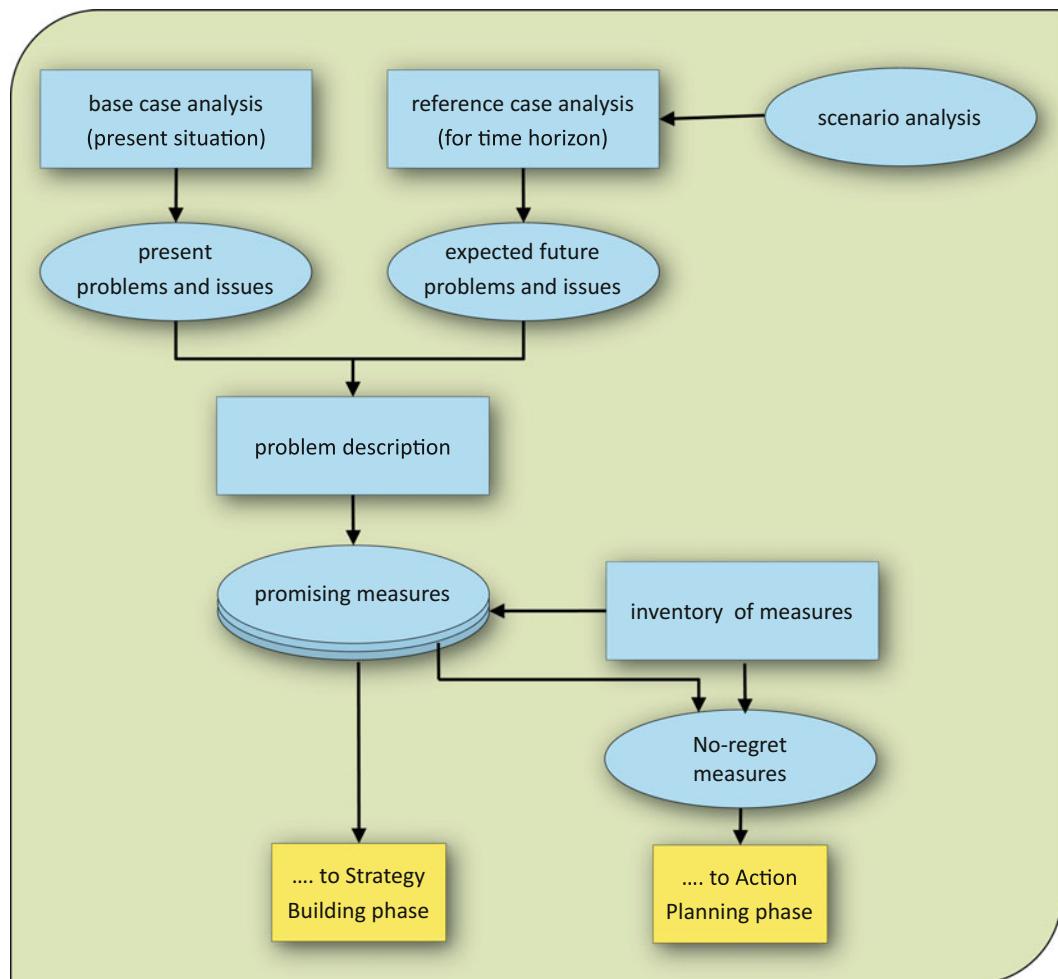
The reference case addresses the future situation by considering the present infrastructure, to which measures are added that have already been decided or are being executed, together with selected scenario conditions. In the reference case an analysis of the performance of the WRS is undertaken if present policies and regulations are continued and followed by the government and the water users.

Problem description—present and future

The problem description should be carried out based on the results obtained from the base and reference case analyses in combination with the problems and issues perceived by the decision-makers and stakeholders. A problem analysis should be expressed as far as possible in terms of the socioeconomic and environmental impacts that have a meaning to the decision makers and stakeholders. An integrated approach is crucial for a solid understanding of the system and its associated problems. The integrated approach can only be achieved if the plan defines the main problems and issues in the basin and its interlinkages. For this, it is important that the plan is aligned with other related plans such as Watershed Plans (erosion), Flood Risk Management (FRM), and Integrated Coastal Zone Management (ICZM), amongst others.

Inventory of potential measures and selection of promising measures

Once the present and future problems are known measures (including "no regrets" that can immediately be implemented) can be identified that will address these problems. An inventory should be made of all the measures that the stakeholders are planning or considering. Based on the quantified problem analysis additional measures might be formulated. The computational framework can be used to determine the impacts of these measures. The most promising measures will be kept for detailed analysis in the next step: Strategy Building.

Fig. 13.9 Structured analysis process

The above described structured analysis process is illustrated in Fig. 13.9.

13.4.2.4   Scenario Analysis

A good plan should not only address the present problems but should also prepare for problems that might arise in future. To predict the future scenario assumptions have to be made. Scenarios are possible developments external to the WRS, i.e., outside the control of the decision makers involved in the project. The most usual scenario components for water resources studies are socioeconomic developments (e.g., growth of population and economic activities) and climate change (including sea-level rise). For the economic evaluation of the plan it might be needed to make assumption about the future prices of energy and food. Changes in diet (e.g., the consumption of more meat) can also be important.

The most used combination of scenario elements are presented in a quadrant of low and high economic growth versus slow and fast climate change. Ideally the whole analysis should be carried out for all kind of scenario combinations and the selection of the best strategy should be based on the evaluation which strategy is able to cope with all these possible future developments. In reality most analyses are carried out for the most likely scenario based on a trend analysis

or Business-As-Usual (BAU). The strategy that follows out of this is then analyzed in a "scenario analysis," to test that strategy on robustness and flexibility for other possible futures. See also Sect. 13.4.3.2 on adaptive management analysis.

13.4.2.5 Quantified Problem Analysis

A problem analysis should address and be expressed in terms of the socioeconomic and environmental or ecosystem impacts that are of interest 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, 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 socioeconomic impacts makes it easier to relate the problems to the (socioeconomic) development objectives that decision-makers have formulated for the particular region or system under consideration.

A good problem analysis will also indicate the measures that can be taken to eliminate, reduce or alleviate the identified problems or to take advantage of new beneficial opportunities. The identification of measures not only helps to clarify the problems and possible solutions; but also helps in the design of the computational framework and the data collection activities. These activities should be designed in such a way that the measures can be evaluated in the analysis phases of the study.

On completion of the initial analysis, project staff (and the decision makers/stakeholders) should have a clear idea about what will be studied in subsequent phases, for what purpose and under what conditions.

13.4.2.6 Identification and Screening of Potential Measures

Once the base and reference cases have been defined, and the problems and bottlenecks identified, measures to address resource management problems can be considered. Measures can be divided into different categories. An inventory of

all possible kinds of actions that can be taken will in general result in hundreds of discrete possibilities. In most cases it will not be practicable to analyze 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 eliminate less attractive or less promising alternatives. It can also be done by using the modeling framework developed for the project but limiting the analysis to a few criteria, such as economic or environmental ones. A third kind of screening analysis is to apply judgment as to criteria effectiveness, efficiency, legitimacy and sustainability. Box 13.3 describes these criteria.

> **Box 13.3. Criteria for screening**
>
> Effectiveness. Measures to be taken are those which solve the most serious problems and have the highest impact on the objectives. Measures to prevent problems will be preferred to those that solve them. Similarly, measures that solve problems will be preferred to those that only control them.
>
> Efficiency. 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 create a law that forces industrial firms to incur the full cost of end-of-pipe wastewater treatment. In Egypt, this would improve the Nile system water quality, and thus improve health of those who drink it and reduce environmental damage. 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 favored by as many stakeholders as possible.
>
> Sustainability. Measures to be taken are those that improve (or at least do not degrade) the present environmental and socioeconomic conditions for future generations.

The aim of the screening process is to identify those measures that should be further analyzed. 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 subbasin. 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 Fig. 13.9.

No regrets

A special category of promising measures are the "no regrets." More realistic we should speak of "likely no regrets" and "low-regret" measures. These are measures on which there is a very large agreement among the decision-makers and stakeholders that these should absolutely be implemented, preferably as soon as possible. It should be ascertained that these measures will not have negative impacts on other measures or will prevent other possible promising measures to be implemented. The reason to define such no regret measures is that in quite some situations there is a huge pressure to actual implement measures and not to wait till (another) big integrated study has been completed and accepted in its full extend. In particular in developing countries there is a big need for proposals for such measures. These measures can proceed immediately to step IV on Action Planning.

13.4.3 Step III—Strategy Building

In the Strategy Building step, 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 the values assigned to the uncertain model parameters is then assessed. Finally, the results of the selected strategies should be presented to the decision-makers. The selection process is depicted in Fig. 13.10.

13.4.3.1 Strategy Design and Impact Assessment
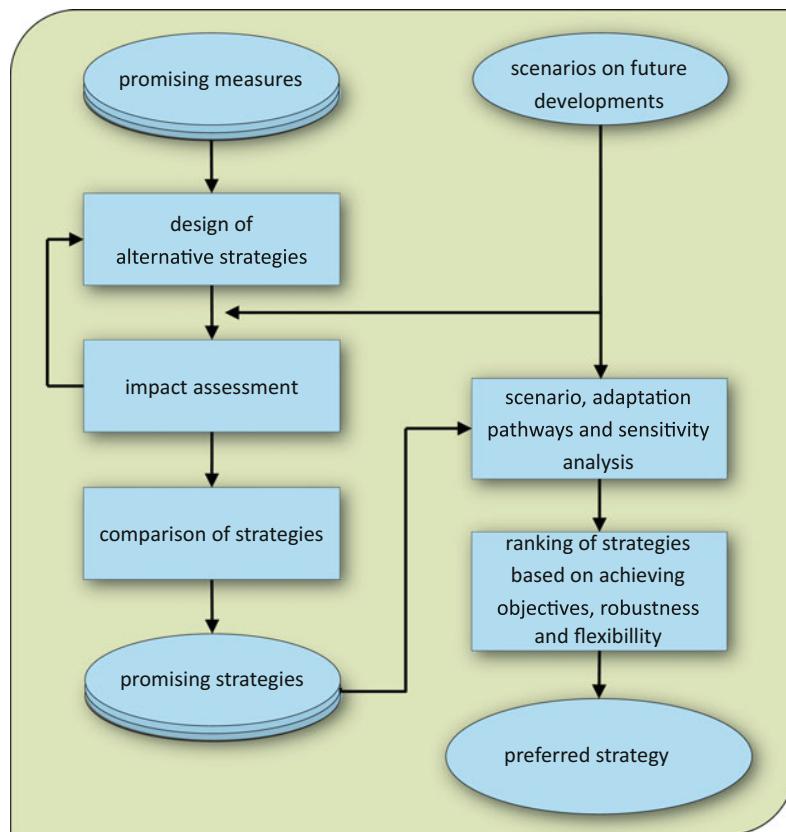
Strategy design involves the development of coherent combinations of promising measures to satisfy the management objectives and meet the management targets if possible. As there are generally many criteria related to these objectives, and probably many expressed in different units, strategy design is not a simple process. Relations among combinations of measures and their scores on the evaluation criteria are complex. The optimum combination may depend on who is asked. Trade-offs among the values of different criteria, and disagreements among various stakeholders, are inevitable.

The design of strategies is an iterative process. One can start by developing strategies on the basis of a single objective such as, for example, reliability of food and energy production or maximum net economic benefits. These strategies define 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.

Evaluation of Alternative Strategies

Strategies can be compared based on their criteria values or scores. 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 going to be important.

It is usually impossible to express all criteria in a single measurement scale such as a monetary value. (We say this recognizing the many attempts to do so by highly respected economists.) Criteria related to environmental quality or ecosystem vitality or the beauty of a scenic view can often be expressed quantitatively but in nonmonetary terms. This should, however, be done in such a way that a ranking is possible on the basis of the chosen 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.

Scenario and Sensitivity Analysis

Before drawing conclusions from planning projects involving uncertain information, and indeed predictions of possible futures, one should analyze the effects of changes in the uncertain assumptions made throughout the analyses. If the selection of a different scenario would significantly change the attractiveness of a selected strategy, then additional study may be required to reduce the uncertainties in that scenario. The sensitivity of the results to changes in model

parameter values and assumptions should be determined and addressed in a similar way.

13.4.3.2 Adaptive Management Analysis
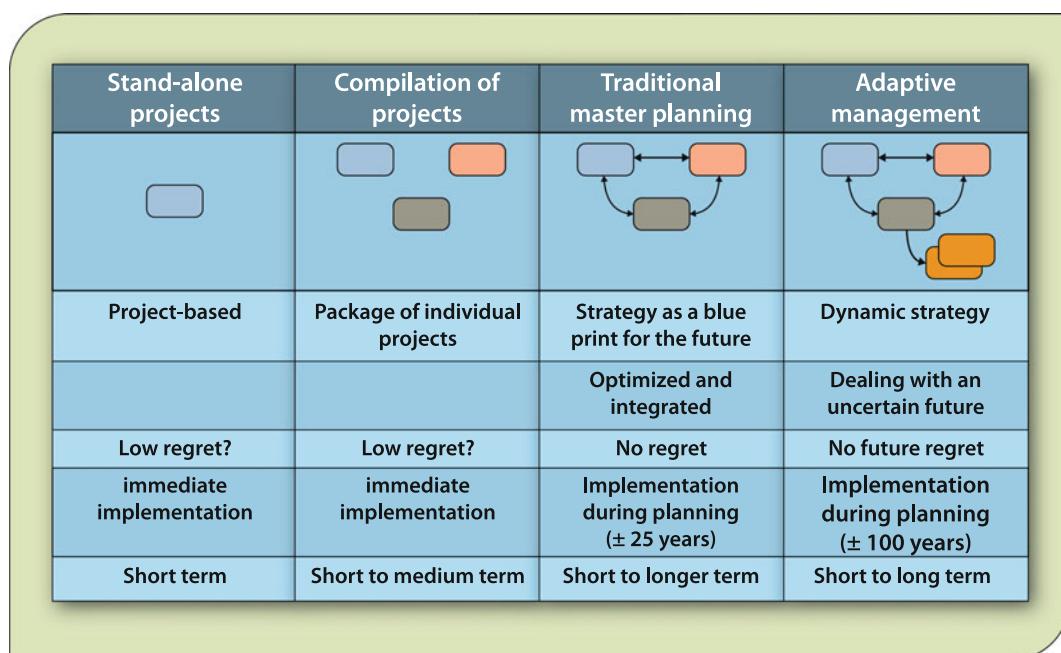
The analysis approach described in the previous section is based on the assumption that it is known what will happen in future. Predictions are made on how population growth, economic growth, spatial developments (e.g., urbanization) and climate change will take place. Some of these developments are quite certain, e.g., population growth for which one can make reasonable good projections. Other developments are much more uncertain such as economic growth and climate change. While we want to be prepared for these future conditions we do not want to run the risk that huge infrastructural investments are being made which later appear to have been overdesigned or even unnecessary.

The way to deal with future uncertainty is to follow an adaptive management approach. An adaptive management approach has to replace the traditional approach of master plans for the basin. The development of implementing stand-alone projects to adaptive management is illustrated in Fig. 13.11.

The message on how to follow an adaptive management approach is given in the right two columns of Fig. 13.11 and is the logical follow-up from the project oriented developments in the two first columns. The figure explains that:

- The project-based approach is straightforward and easy to implement. This approach does not consider the (positive and negative) interaction of the project with other projects.
- The interaction is taken into account when related projects are considered in a package of projects. However, the overall system is not integrated yet and not optimized.
- The traditional master planning tries to optimize the overall system. The projects are implemented as components of an integrated strategy. The implementation of the strategy includes an optimization of the various projects over the planning period which is usually between 15 and 30 years, for which a cost–benefit analysis usually applies. Such a

Stand-alone projects	Compilation of projects	Traditional master planning	Adaptive management
Project-based	Package of individual projects	Strategy as a blue print for the future	Dynamic strategy
		Optimized and integrated	Dealing with an uncertain future
Low regret?	Low regret?	No regret	No future regret
immediate implementation	immediate implementation	Implementation during planning (± 25 years)	Implementation during planning (± 100 years)
Short term	Short to medium term	Short to longer term	Short to long term

Fig. 13.11 Planning approaches in water resources management

master planning approach does not consider the long-term uncertainties that are involved in socioeconomic developments and climate change. If the predicted changes in socioeconomic conditions and climate do not materialize this might lead to "future regret."

- To reduce future regret a planning period of up to 50 or even 100 years needs to be considered. As the lifetime of most structural measures (dikes, floodways, reservoirs, etc.) are designed for a period of 50–100 years, it is wise to incorporate future uncertainties in boundary conditions in their designs and make them part of a dynamic strategy. The adaptive approach not only tells us what to do now but also gives directions on what to do when the conditions develop differently.
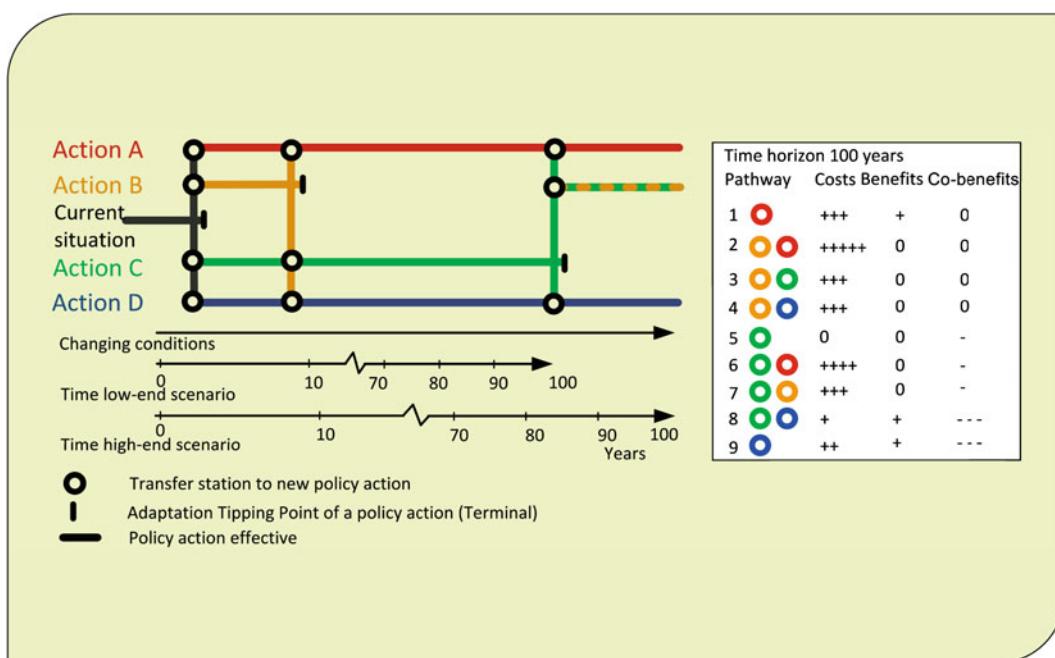
Adaptive pathways

Various methods have been developed that enable us to deal with future uncertainties. Recent methods include Decision Trees (Ray and Brown 2015) and Dynamic Adaptation Policy Pathways (DAPP; Haasnoot et al. 2013). The Decision Trees is a repeatable method for evaluation of

climate change risks to new development projects. DAPP identifies tipping points that determine in time when a certain policy or action is no longer acceptable and (another) action is needed. By exploring all possible actions you can develop adaptation pathways that will minimize the regret. The Adaptive Pathway Approach is illustrated in Fig. 13.12. The approach requires that many conditions are explored (pathways, scenarios, long time series). For that reason the models used in an adaptive pathway analysis are sometimes limited versions (meta-models) of the ones described in this book. See Haasnoot et al. (2014).

Following an adaptive pathways approach basically means that two additional criteria should be considered in decision-making:

- **Robustness**: how robust is the existing strategy when the future develops differently than expected? Will the strategy then still achieve the objectives?
- **Flexibility**: how changeable is the strategy when it appears that the future develops differently than expected and we need to change the strategy?

Fig. 13.12 Adaptive pathways approach

Robustness and flexibility often have a strong relationship with costs. A robust strategy can be more costly (big reservoirs, high dikes, etc.). A flexible strategy (many small reservoirs, build in time) can also appear to be more expensive in the end. These costs need to be taken into account when deciding on a strategy.

13.4.3.3   Presentation of Results—Preferred Strategy

Presentation of the selected promising strategies to decision-makers may be by means of briefings, presentations, and summary reports among other means. The level of detail and the way project results are presented should give an overview of the results at an appropriate level of detail for the audience involved. Visual aids such as score cards and interactive computer presentations of study results are often very helpful for promoting 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 Table 13.2. Scorecards can contain numbers only, or the relative value of the criteria can be expressed by plusses and minuses, or a color or shading. The purpose of scorecard presentations is to present a visual picture of the relative attractiveness of the alternatives based on various criteria. Scorecards can also help viewers 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 he considers most appropriate.

13.4.4   Steps IV and V—Action Planning and Implementation

Once the preferred strategy has been selected this strategy should be translated into concrete actions. Careful planning and coordination is

required as many authorities will be involved in the implementation. The action plan will have an "open" and "rolling" character, meaning that it is not static or prescriptive, and leaves room for individual decision-makers to further elaborate upon in relation to their own responsibilities. On the other hand, the action plan should be concrete, by assigning clear responsibilities for carrying out the activities involved. It also should include the budgetary requirements for the implementation, including investments and recurrent costs.

13.4.4.1   Investment and Action Plan

The action plan translates the selected strategy in concrete actions. For each of these actions it should be clear:

- what: concrete actions that have to be carried out for each of the measures included in the strategy to get it implemented?
- who: the prime decision-maker/stakeholder responsible for carrying out the action and who will take the lead in the implementation;
- how: the steps to be taken and the consultative process involved;
- when: the time planning; and
- financing: where will the money to implement the action come from?

What

An integrated planning analysis is usually carried out at pre-feasibility level. A rough description of the measures will been included in the strategy and the assessment was based on first estimates of costs and benefits. Depending on the type of measure, feasibility studies should be completed before the measures can actually be implemented. Often these feasibility studies are combined with detailed (technical) design of the measures.

Who and How

The Action Plan aims to stimulate the coordinated development and management of the water resources. This is illustrated in Fig. 13.13, which presents the Implementation Plan for water resource development in Central Cebu in the

Fig. 13.13 Implementation plan (taken from Cebu study)

The table below summarizes the readable measure information from Fig. 13.13 (institution-involvement markers are indicated in the original matrix with symbols ● = responsible, o = involved/consulted, x = affected, ? = uncertain):

Category	Sub-category	Recommended Measure	Phase	No.	Type	Total investment (M Pesos)	Recurrent costs (M Pesos/y)	Starting year effective operation	Needed lead time (prep. + constr.)
Develop more resources	West	Groundwater wells West	I	9	invest.	849	38	2006	-
	Uplands	Spring boxes + small impoundments + trucking for drought periods	I	8	invest.	737	126	2006	-
	East, non-MCWD area	Groundwater wells East	I	9	invest.	639	28	2006	-
	East-MCWD area, excl. Mactan	Groundwater production wells by MCWD	I	9	invest.	588	26	2006	-
		Groundwater wells MCWD area by private initiatives	I	9	invest.	588	26	2006	-
		Luyang-Carmen weir	I	4	invest.	3.169	124	2016	2+2
		Northern well fields: Liloan, Compostella, Kotkot, Danao, Carmen, Luyang (excl. Luyang weir)	I	9	invest.	3.169	124	2006	2+2
		Lusaran Dam	I	2	invest.	2.536	77	2027	6+2
		Kokot Dam	I	3	invest.	905	59	2022	6+2
		Southern well fields: Napo/Carcar river, Pangdan, Minglanilla	I	10	invest.	2.490	86	2014	2+2
	Mactan	Horizontal wells	I	11	invest.	51	5	2006	-
		Shallow fresh water wells	I	11	invest.	276	56	2008	-
		Shallow brackish water wells + reverse osmosis	I	13	invest.	80	42	2006	-
		Desalination by MCWD for industry	I	13	invest.	297	55	2010	2+2
		Desalination by MCWD for domestic use	I	13	invest.	1.393	167	2012	2+2
		Desalination by industry (private initiative)	I	13	invest.	297	55	2010	2+2
Water demand reduction	Reducing losses	Rainfall harvesting: urban, rural, industry	I	12	invest.	515	-	2006	-
		Leakage reduction / rehabilitation distribution system MCWD	I	14	invest.	820	41	2007	-
		Adequate and differentiated water pricing	I/II	20	study	pm	pm	2007	1
		Promotion of water saving equipment and less consuming production	II	21	man.	-	-	-	-
	Awareness	Awareness raising high demand Cebu neighbourhoods	I	22	man.	0	2,6	2006	-
		Awareness raising - general	I	23	man.	0	6,9	2006	-
Protecting the resources	Watershed management	Land use practices / watershed management	I/II	30	reg.	-	-	-	-
		Gabion dams	III	31	invest.	-	-	-	-
	Water quality improvement	Improved solid waste management	II	34	man.	-	-	-	-
		Well head protection - spatial planning recharge areas	II	35	reg.	-	-	-	-
		Prevent sand and gravel mining in rivers (enforcement)	II	36	reg.	-	-	-	-
		Implement strict effluent permitting (EMB)	II	37	reg.	-	-	-	-
		Urban sewage systems in building requirements	I	38	reg.	0	0	2007	1
		Development of urban sewage systems and treatment	II	39	invest.	-	-	2010	3
		Sanitary programs in uplands	III	40	invest.	-	-	-	-
Water governance	Organisational	Development institutional setting for IWRM (integration and coordination - Board, TS etc.)	I	50	instit.	pm	pm	2006	0
	Regulation and control	Family planning and migration control (env. related)	I	54	man.	pm	pm	-	-
		Water withdrawal quantity and quality control	II	52	man.	pm	pm	2007	0
		Implement and enforce spatial planning (urban, industrial, etc.)	II	55	man.	pm	pm	-	-
		Priority rules (allocation) during dry periods	I	56	reg.	pm	pm	-	-
Research and development		Improve data collection, analysis and presentation, improve decision support tools	II	61	capac.	pm	pm	2007	-

Philippines. The measures included in the plan will involve or affect many stakeholders. All these stakeholders (based on the outcomes of the stakeholder analysis and designed participatory planning process) should therefore be included in some way in the implementation process in order to guarantee a successful implementation and a sustainable benefit of the particular measure. In general the following roles can be distinguished:

- Responsible: the stakeholder has the first responsibility for the implementation of the measure but will co-operate with and/or consult other stakeholders in this process. In

Fig. 13.13 this is indicated by the symbol: "●".

- Co-operate: the stakeholder has an important say in the implementation of the measure but is not the first responsible and is expected to work with other stakeholders in this matter. In the figure this is indicated by the symbol: "○".
- Consult: the stakeholder has an interest in the implementation of the measure and will be consulted by the first responsible. In certain cases permission will be needed before the implementation can take place. In the figure this is indicated by the symbol: "x".

When

The action plan should also specify the timing of the implementation. When will (the preparation of) the implementation start, and when should the implementation be finalized. This information is needed for the overall investment plan but also because some measures will depend on the completion of other measures.

13.4.4.2 Financing—Investment Plan

An important, if not the most important, part of the Action Plan is to determine how the action will be financed. The sources of the financing will largely depend on the type and size of the measure. As water resources management is mainly a governmental task, most of the finances will come from public sources. These can be from the national budget (possibly supported by donor funds) or from local (province, municipality) budgets. In some cases private funding can be considered in PPP (Public Private Participation) constructions. This seems in particular attractive when there is a good possibility for payment by the stakeholders of the services that will be provided. Examples where PPPs can be considered are urban public water supply and hydropower production.

The investment plan should also address how the recurrent costs (operation and maintenance) of the implemented projects will be recovered. Preferably this should be done based on fees to be paid by the people that benefit from the project.
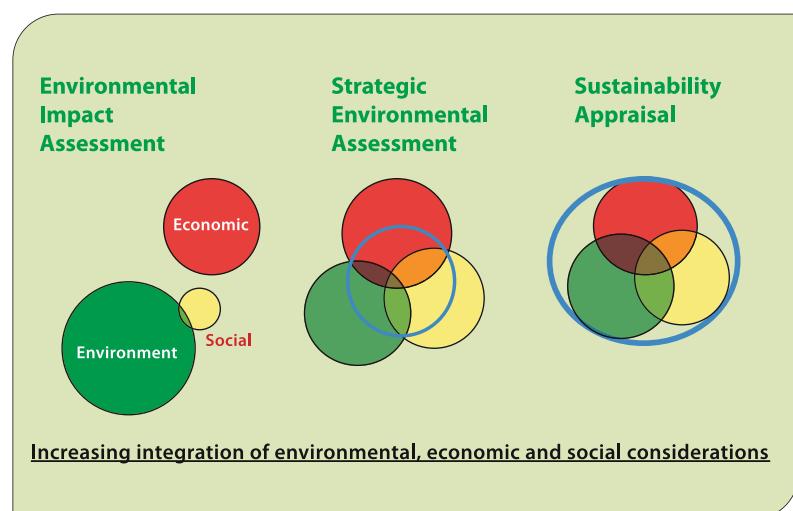
13.4.4.3 Feasibility Studies and Environmental Impact Assessment

A feasibility study should include a more detailed study of the projects (measures) proposed in the plan. Commonly a feasibility study includes some 5 areas of feasibility:

- technical
- social/environmental
- political/legal
- financial/ economic
- operational and scheduling

A feasibility study for a good implementation planning will often include a more detailed assessment of the possible socioeconomic and environmental impacts of some of the measures

Fig. 13.14 Applying SEA (OECD 2006)

that comprise the preferred strategy. There are several types of assessment depending on the focus of the study. As depicted in Fig. 13.14 the most well-known are: Environmental Impact Assessment (EIA, for infrastructure projects), Strategic Environmental Assessment (SEA, mainly used in policy development) and Sustainability Appraisal (SA).

13.4.4.4   Promotion

After the action plan has been established one needs to find ways to increase the influence of stakeholder groups that favor the implementation of the action but lack influence; to change the attitude of influential groups that are opposing this action; and to use the positive attitude of influential groups that are in favor of this action. The results of the stakeholder analysis are used for the identification of the stakeholder groups. As illustrated, the matrix highlights the strategy toward project acceptability or appreciation and therefore smooth implementation.
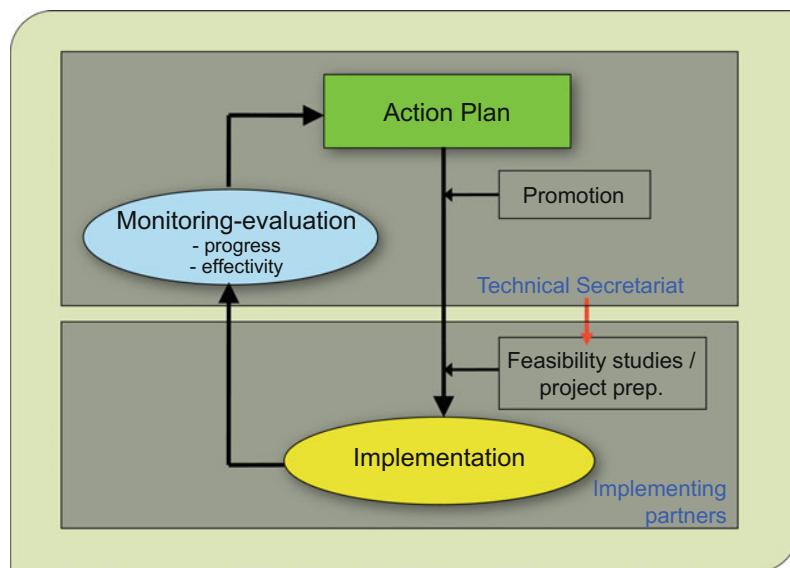
To create maximum awareness, enthusiasm and support for selected projects within the Action Plan the selected stakeholder groups need to be provided with the right information on the project. Additionally, involving a selection of stakeholders in project preparation and implementation will assist in making them enthusiastic about the project. To do this effectively, a mix of marketing options can be used. Appropriate marketing options might be:

- mass one-way communication for the general public (such as newspapers, radio, television plus more traditional media in the more rural areas);
- selective one-way communication for selected stakeholders groups (direct mail, brochures with more specific information dedicated for the selected group); and
- personal two-way communication between the project promoter and selected stakeholders groups (education method, outreach method or more risky word-of-mouth method).

13.4.4.5   Monitoring and Evaluation

An overview of the implementation framework is given in Fig. 13.15. This implementation framework applies for both Steps IV (Action Planning) and V (Implementation). The actual implementation of most of the measures will take place by decentralized agencies of national ministries or at local governmental level and their related utilities, districts, and associations. Where

Fig. 13.15 Implementation framework

needed feasibility and engineering studies will be carried out before the actual implementation and/or construction can take place.

Above the implementation level there should be a guidance and coordination level, e.g., a Technical Secretariat (TS) at basin level. Periodically a monitoring report compiled by the TS can track the progress made in implementing the measures of the Action Plan and the effectiveness of these measures in meeting their objectives. Insufficient progress may lead to an adjustment of the Action Plan. The TS may also provide assistance to the implementing partners, e.g. the local government agencies, as they carry out feasibility studies. The TS should be able to support them by providing data and possibly other relevant information from their Management Information System (MIS).

13.5   Making It Work

The framework of analysis presented in Fig. 13.3 includes next to the five steps of analysis two crucial blocks that play a role in several of these steps and deserve special attention. The first one is the stakeholder engagement in the analysis. Involving stakeholders and making sure that their ideas and suggestions are taken into account is an absolute requirement to develop a consensus and support for the ultimate plan that is to be implemented. There is no guarantee that a consensus will be reached, however. Involving stakeholders in each stage of the planning framework takes extra time and money, but if any ultimate plan is to be accepted and prove sustainable, there is no other choice. At a minimum, any plan that is derived from this process should be an informed one, based on inputs from all affected stakeholders and decision makers.

13.5.1   Stakeholder Engagement

The stakeholders that should be involved in a planning process will depend on the specific basin that is being addressed. In general the stakeholders will be all people and/or organizations that:
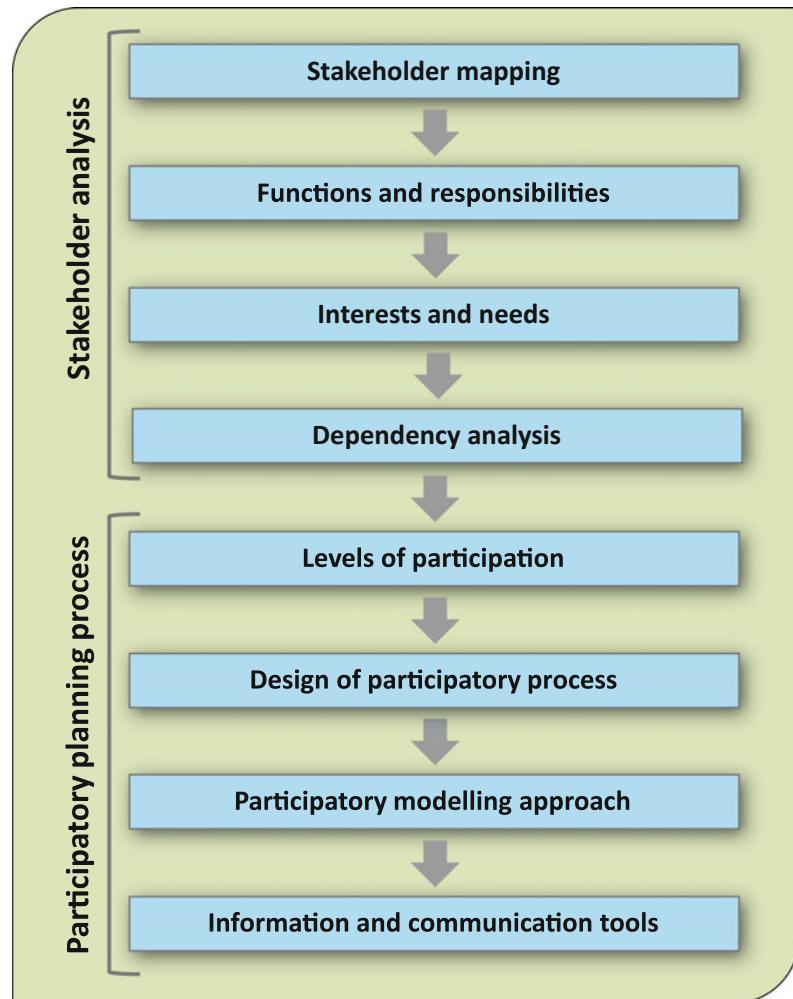
- will be effected by the plan; and
- are needed to implement the plan.

An integrated plan and its implementation depend to a large extend on the acceptance and ownership of the plan by the decision makers and stakeholders at national and basin levels. A participatory planning process is therefore indispensable for sustainable WRM. A participatory planning process is the results of a set of steps, as depicted in Fig. 13.16. However, the order of the steps can vary according to the local situation and conditions. The prerequisite for the design of a participatory planning process is a good stakeholder analysis. The stakeholder analysis is a supporting planning tool that supports the identification of stakeholders and its engagement. Particularly, this analysis technique supports the task of identifying and in some occasions classifying the stakeholders according to their functions, capacities, interests, concerns and needs, as well as their dependencies (including power relations among them).

Based on the results of the stakeholder analysis the participatory planning process is defined. First, it is crucial to define the levels of participation of the various stakeholders. The level of participation of each group of stakeholders varies depending on the stakeholder analysis and on the maximum level of participation that the client of the study wants to achieve. The second step is the design of the participatory process. This will be adapted to the agreed levels of participation and stakeholders involved. The design of the participatory process needs to take into account the modeling approach (informed decision making) so it is carried out in a participatory manner (step 3). Finally, as illustrated in Fig. 13.16, the design of the participatory planning process needs to consider the information and communication tools used for disseminating and communicating the information to the various groups of stakeholders as illustrated in the power-interest matrix of Fig. 13.17.

Fig. 13.16 Steps in a stakeholder analysis and participatory planning process

Stakeholder analysis
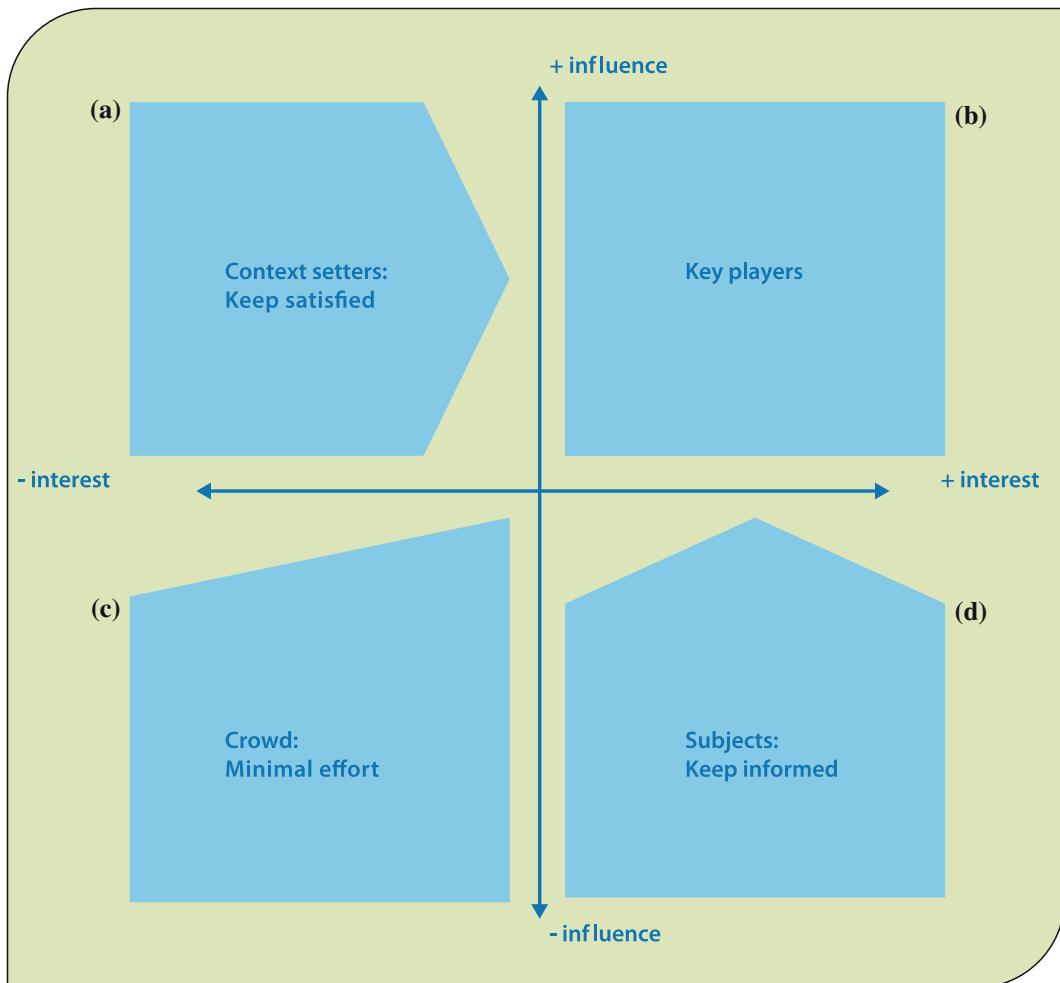
A stakeholder analysis provides a better understanding of the perceptions, concerns, roles, interests, and needs of the stakeholders and contributes to a better approach to the solution. It also helps reduce the possibility of forgetting important risks. Finally, this technique increases the chance that the various groups of stakeholders are willing to cooperate in solving the identified problems and issues.

A good stakeholder analysis should contain at least the following steps:

(1) Situation analysis as point of departure.
(2) Inventory of the stakeholders involved (e.g., primary, secondary and tertiary stakeholders).
(3) Mapping of formal relations according to their functions and responsibilities.
(4) Inventory of interests, perceptions, and needs.
(5) Mapping of interdependencies.

Levels of Participation

The various stakeholders are grouped into the different levels of participation according to the
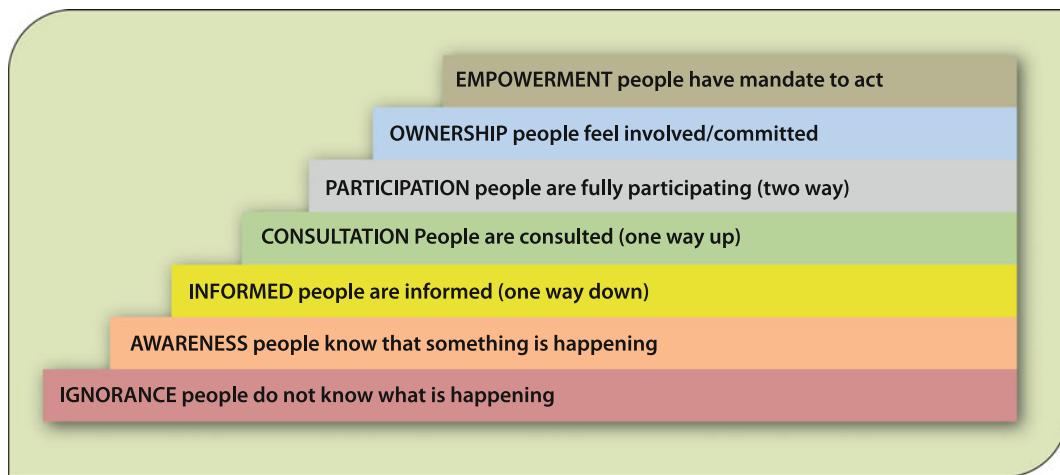
Fig. 13.17 Power-interest matrix

outcomes of the stakeholder analysis, as illustrated in Fig. 13.18:

- Ignorance: where a stakeholder is not aware of what is happening;
- Awareness: where a stakeholder is aware that something is happening;
- Informed: where a stakeholder has been specifically provided with information and is left to decide what to do with it. The emphasis is on the one-way provision of information, with no formal option for the stakeholder to provide feedback, negotiate, or participate in the decision-making process;

- Consultation: where a stakeholder is asked to provide information inputs to the planning process. Information flows are likewise one-way, but in the opposite direction. That is, information is extracted from stakeholders although no commitment is given to use it;
- Discussion: at this level are fully participating and are asked to give advice and recommendations. Here information flows in both directions between stakeholders operating with different interests and levels of influence, and also between these stakeholders and the organizing team (technical team). Since two-way interactions occur, there is room for

Fig. 13.18   Levels of participation

alternative ideas, solutions and/or strategies to emerge;

- Co-Design: at this level stakeholders are actively involved in problem analysis and problem design, which fosters ownership, but where final decision-making powers reside with the governing agencies;
- Co-Decision-Making: here decision making powers are shared with those participating stakeholders, leading to their empowerment with respect to the policy/planning decision taken. Typically decisions in these contexts would emerge from a process of stakeholder negotiation.

The first levels (from Ignorance to Consultation) could be thought of as top-down management/planning approaches toward participation, where stakeholders have little control over the decision-making process. The final three levels are more appropriately considered as bottom-up approaches toward participation where stakeholders are much more active and have much more control over the decision-making process.
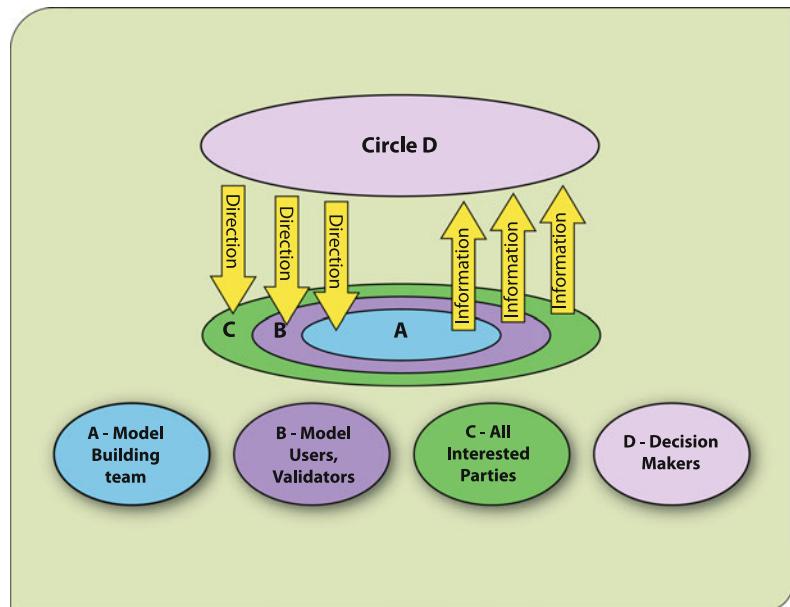
Design of the participatory planning process

The design of the participatory planning process needs to take into consideration the River Basin planning framework and the data and modeling tools used. Participatory planning tools and techniques enable participants (stakeholders) to influence development initiatives and decisions affecting them. The tools promote sharing of knowledge, building up commitment to the process and empower the group to develop sustainable strategies.

The participatory and informed planning process makes use of the "Circles of Influence" model (Fig. 13.19) that enables to structure participation to limit numbers but not the influence of specific groups of stakeholders (Cardwell et al. 2008; Bourget 2011). Under this model trust is developed in concentric circles; planners and managers work to develop trust with leaders and organizations that other stakeholders already trust. That is, those most directly involved in policy analysis activities (i.e., planners, managers, and modelers who do most of the actual work; Circle A) who communicate with trusted leaders and major stakeholder representatives at the next level (Circle B). These stakeholders then in turn provide a trusted link to all other interested parties, who have much less direct involvement (Circle C). Ideally, Circle B participants would be active in professional or issues-oriented organizations and provide links to others whose interests they represent. Hence, Circle C stakeholders should see their interests represented in Circle B, and have formal

Fig. 13.19 Participatory planning structure based on circles of influence (*source* Cardwell et al. 2008)

opportunities to shape the work of Circles A and B via these representatives. The levels of involvement of those stakeholders in Circle C can vary from Consultation to Awareness. A fourth circle (Circle D) includes decision makers such as agency heads and elected officials, who have been given the authority to accept or reject the recommendations of the policy analysis. For a good participatory and informed planning process it should be clearly identified and engaged throughout the planning process with direction and information flows possible to and from all circles.

Other aspects to be considered for the design of the participatory planning process are:

- Timing of stakeholder involvement. This will be dependent on the Circles of Influence and levels of participation.
- Stakeholder participation in the modeling process (Participatory Modeling). Mainly those stakeholders in the Circles A and B will be regularly involved in some of the phases of the modeling process. The involvement can be concentrated in (i) early and later stages of

the modeling process, (ii) construction of the model, (iii) some of the activities prior to model construction, or (iv) only after the final model has been built.
- Type of stakeholder involvement. This can be either individually, with homogeneous (stakeholders with similar interests and problem perceptions) or heterogeneous groups.
- Information and communication tools. Information dissemination (e.g. face-to-face workshops or online platforms) and communication tools need to be adapted to the background conditions of the various groups of stakeholders. This is particularly important for participatory model construction and use, as well as, for the promotion of the plan. The selected marketing options for creating awareness, enthusiasm and support for selected projects within the action plan by stakeholders will vary depending on the results of the stakeholder analysis (Fig. 13.17) and levels of stakeholder involvement (Fig. 13.18). For more information about plan promotion see Sect. 13.4.4.3.

13.5.2 Using Models in a Planning Process

13.5.2.1 Managing Modeling Projects

There are some steps that, if followed in modeling projects, can help reduce potential problems and lead to more effective outcomes. These steps are illustrated in Fig. 13.20. Some of the steps illustrated in Fig. 13.20 may not be relevant in particular modeling projects and if so, these parts of the process can be skipped. Each of these modeling project steps is discussed in the next several sections.

Creating a Model Journal

One common problem of modeling projects once they are underway occurs when one wishes to go back over a series of simulation results to see

Fig. 13.20 The modeling project process is typically an iterative procedure involving specific steps or tasks

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, project documentation is often neglected under the pressure of time and perhaps because it is not as interesting as running the models themselves.
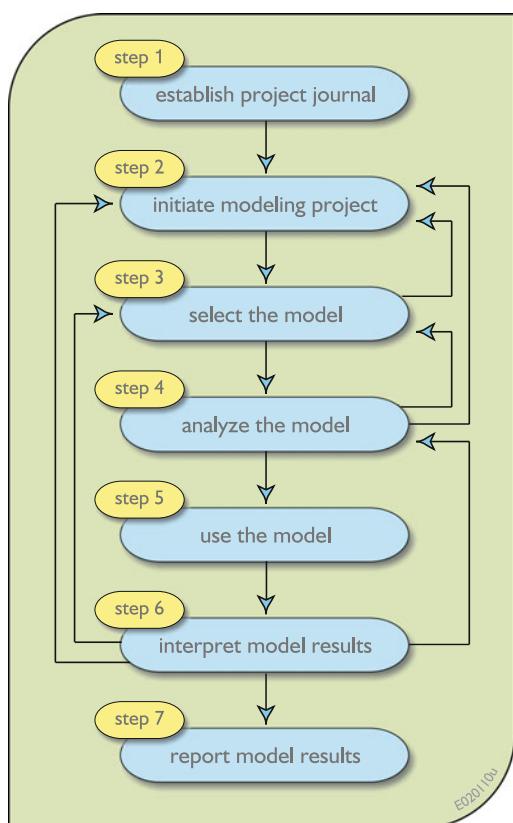
Initiating the Modeling Project

Project initiation involves defining the problem to be modeled and the objectives that are to be accomplished. 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 modeler. In addition, problem perceptions and modeling objectives can change over the duration of a modeling 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 modeling approach, or even modeling in general, is the best way to obtain the needed information. What are the alternatives to modeling or a particular modeling approach?

The objective of any modeling 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 modeled. 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 modeling project.

Proposed modeling activities may have to be justified and agreements made where applicable. Any client at any time may wish for some justification of the modeling 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 modeling 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 modeling project outputs (data files) and reports.

Selecting the Model

The selection of an existing model to be used in any project, as opposed to developing a new one, depends in part on the processes that will be modeled (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 and use. These boundaries define 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 modeler'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 if help is 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:

- 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.
- Consider model and computational cost. Today computing costs are 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.

Analyzing the Model

Once a modeling approach or a particular model has been selected, its strengths and limitations should be assessed. 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 Chap. 8). 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 modeling 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 sufficiently "correct". There is no way to prove it. ["All models are wrong but some are useful" Box (1976).]

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 modeling 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 modeling 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.

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 modeler 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?
- Were 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?

Some of these questions may not apply, but 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.

Interpreting Model Results

Interpreting the information resulting from simulation models is a crucial step in a modeling

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 modeling project objectives.

The results of any water resources modeling 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.

Reporting Model Results

Although the results of a model should not be the sole basis for policy decisions, modelers have a responsibility to translate their model results into policy recommendations. Policymakers, 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 policymakers. 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.

13.5.2.2   Evaluating Modeling 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 modeling.

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 modeling 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 modeling has been made and implemented. The extent to which the information and understanding resulting from modeling 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). Modeling 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 the adverse ecological consequences of converting rivers to canals, and whether or not anyone will care. 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 modelers 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 analyzed and implemented modeling 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.

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 data-driven 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, and we think will continue to have, a major impact on the state of the practice in using models in support of water resources planning and management activities.

13.6 Conclusions

The effectiveness of strategies for dealing with issues of water quantity and quality, and their variability, has a major impact on the well-being of living species, and even the survival of some. How well water is managed also impacts the functioning and resilience of ecosystems, the vitality of

societies, and the strength and growth of economies. Fortunately we humans can determine which water resources development and management strategy will work best in a given situation, not only for the immediate future but in the long-run as well. And if conditions change, our strategies can adapt. To accomplish this we need to identify and evaluate the effectiveness of the water resources development and management alternatives available to us in an economic, hydrologic and sociopolitical environment that seems to be a constantly changing. We can do this through the use of various models, developing preferred strategies based in part on their results, and informed by the concerns and objectives of stakeholders and the decision making institutions.

This book has focused on ways of developing and using various optimization and simulation modeling methods for analyzing and evaluating water resource development and management alternatives. This final chapter has presented some guidelines for carrying out water resources planning projects, including its modeling components. Such projects are typically very complex and challenging.

Water management planning projects must address a complex and interconnected web of science, engineered infrastructure, legal regulations governing water use, societal expectations, and institutional structures and authorities that have evolved over time. Much of the current complexity that exists in various regions of the world has developed over time in response to changing interests and objectives of water users and environmental considerations. Although the impacts of changes in the climate on water supplies and demands are generally recognized, these ongoing changes as well as the linkages between environmental and societal factors in specific basins and regions all lead to major uncertainties in the future.

The guidelines discussed in this chapter have been developed and used by Dutch experts in Deltares to assess water resources systems and to develop plans and strategies for managing them. Deltares has been actively involved in numerous water resources planning and management projects throughout the world. The approach described in this chapter illustrates how these projects are conducted, and the major factors that are considered while conducting them. The effects and impacts of some of their projects have been relatively local and required consideration of only a few sectors of the economy. Other, more comprehensive projects have had national or international impacts, and led to transboundary (international) compacts.

Clearly each water resources system is unique with respect to its management issues and problems and its institutional environment. Project planning and analysis approaches must adapt to these situations. Hence, each project will differ, and will no doubt need to deviate from the suggested guidelines presented in this chapter. Other approaches are possible and may be equally effective. What remains important in all cases is the establishment of a comprehensive, systematic process of planning and analysis together with constant communication among planners, decision-makers and the interested and affected public. The end result should be an improved, more sustainable, and equitable water resources development plan and management policy, appropriate for the region and its people.

Box 13.4. 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
	  – Number of permanent jobs (#)
	  – Number of temp. jobs (mn-year)

2. Increase income of people
– Improve income position of farmers
– Improve equity in income distribution
	• Farmer net income (Rp/year)
	• Difference in benefits of WRM strategies per capita between:
	  – Kabupatens (%)
	  – Urban/rural areas (%)
	  – Income groups (%)

3. Increase the non-oil export production (shrimps, tea, and rubber)	– Export value (Rp/year)
4. Support economic development in an economically efficient way	– Total annual. benefits (Rp/year) – Total annualized costs (Rp/year) – B/C ratio (–) – IRR (%) – NPV (Rp/year) – Total capital required (Rp) – Foreign currency required (%) – Total construction costs (Rp) – Total O&M costs (Rp) – Sectoral value added (Rp/year) – GRP (Rp/year)
User-related (sectoral) objectives and criteria	
1. Increase agricultural production (3% per year)	– Padi (ton/year) – Palawija (ton/year) – Export value of crops (or import substitution) (Rp/year) – Unit costs water supply (Rp/m^3) – % failure meeting demand (%)
2. Increase power production (–)	– Installed capacity (MW) – Power production (GWh/year) – Failure meeting firm power (%) – Price of power prod. (Rp/Kwh) – Energy production value (Rp)
3. Increase fish production (–)	– Fish produced (ton/year) – Fish pond area (ha) – Export value (Rp/year)

4. Support industrial development
  • Water supply for industry (full supply)
  • Provision of opportunity for discharge of waste water
	– Amount of supply (m^3/s)
	– Cost of water supply (Rp/year)
	– Unit costs water supply (Rp/m^3)
	– % failure meeting demand (%)
	– Cost to maintain water quality standards (Rp/year)

5. Enhance water-related recreation	
Environmental and public health related objectives and criteria	

1. Improve public health
  • Improve drinking water supply
    urban: BNA, IKK and major city programs:
60 l/cap/day, serving 70 %
    rural: 55 %
  • improve flushing
    (1 L/s/ha in urban area)
	– Supply (1/day/ capital)
	– % of people connected (Rp/m^3)
	– Price of drinking water (%)
	– % failure meeting demand (%)
	– Volume of flushing water (m^3/s)
	– Unit costs (Rp/m^3)
	– % failure meeting demand (%)

Objectives	Evaluation criteria
2. Improve/conserve natural resources and environment ・Erosion and sedimentation control (erosion <1 mm/year) ・Conservation of nature ・Water quality	– Area severely eroding (ha) – Erosion (mm/year) – Sediment yield (tons/year) – Reafforestated area (ha) – Replanted area (ha) – Terraced area (ha) – % external wood supply to total wood demand (%) – Concentration water quality parameters (ppm)
3. Provide flood protection (return period: depending on value of endangered area)	- return period [years] - flood alleviation benefits (reduced damage) [Rp/year] - flood control cost [Rp/year] - number of people in endangered areas [#] - 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. Takes care of maximum agreement of proposed strategy with existing competence and responsibilities of agencies concerned	– Deviations from/conflicts with existing competence and responsibilities

[a]Kabupaten = Indonesian administrative unit
[b]Padi = Rice crop
[c]Paliwija = Non-rice crop
[d]Rp = Rupiah

Box 13.5 Example 2: Score-card 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**	Mfeddan	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	2812	2075	2153
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

	Unit	1997 base	2017 reference case	Strategy facing the challenge
Industry: costs 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 tot. 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 farmers	LE/year	5362	4629	4309
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 Nile system Outflow to sinks from Nile system	BCM	16.3	17.6	12.5
Overall water use efficiency 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 Nile system	m^3/feddan/yr	4495	3285	3866
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

	Unit	1997 base	2017 reference case	Strategy facing the challenge
Water quality shallow groundwater	−, 0, +	0	−	−
Ecology and sustainability Sustainability: use of deep groundw.	Abstr/pot	0.15	1.00	1.00
Condition Bardawil (Ramsar site)	−, 0, +	+	−	+
Condition coastal lakes	−, 0, +	0	−	0

[a]UFW = Unaccounted for water (the water that is lost in the system)
[b]feddan = 0.42 ha
[c]LE = Egyptian pound

References

Bourget, L. (Ed.). (2011). *Converging waters: Integrating collaborative modeling with participatory processes to make water resources decisions*. Institute for Water Resources, U.S. Army Corps of Engineers.

Box, G. (1976). Science and statistics. *Journal of the American Statistical Association, 71*, 791–799. doi:10.1080/01621459.1976.10480949.

Cardwell, H., Langsdale, S., & Stephenson, K. (2008). *The shared vision planning primer: How to incorporate computer aided dispute resolution in water resources planning*. Alexandria, VA: Institute for Water Resources. IWR Report 08-R-02 (2008).

Goeller, B. F. (1988). *A framework for evaluating success in systems analysis*. Report P-7454. Santa Monica: Rand Corporation.

GWP. (2000). Integrated water management, TEC Background Paper no. 4. Stockholm: GWP.

Haasnoot, M., Van Deursen, W. P. A., Guillaume, J. H. A., Kwakkel, J. H., Van Beek, E., & Middelkoop, H. (2014). Fit for purpose? Building and evaluating a fast, integrated model for exploring water policy pathways. *Environmental Modelling & Software, 60*, 99–120. ISSN 1364-8152.

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & Maat, J. T. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change, 23*(2), 485–498.

OECD. (2006). Applying strategic environmental assessment—Good practice guidance for development co-operation. www.oecd-ilibrary.org/development/applying-strategic-environmental-assessment_9789264026582-en

Ray, P. A., & Brown, C. M. (2015). *Confronting climate uncertainty in water resources planning and project design, the decision tree framework*. Washington, United States: World Bank.

Additional References (Further Reading)

ADB, GIWP, UNESCO, WWF. (2013). River basin planning, principles, procedures and approaches for strategic basin planning.

Asian Water Development Outlook. (2013). Measuring water security in Asia and the Pacific.

Bousset, J.-P., Macombe, C., & Taverne, M. (2005). *Participatory methods, guidelines and good practice guidance to be applied throughout the project to enhance problem definition, co-learning, synthesis and dissemination*. SEAMLESS.

Common Implementation Strategy for the Water Framework Directive. (2000/60/EC). Guidance document 8 on public participation in relation to the water framework directive. http://ec.europa.eu/environment/water/water-framework/facts_figures/guidance_docs_en.htm

GWP. (2005). *Integrated water resources management plans—Training manual and operational guide*. Stockholm: GWP.

GWP Toolbox www.gwp.org/ToolBox/

GWP, INBO. (2009). A handbook for IWRM in basins. Stockholm/Paris: GWP/INBO.

NSMC. (2010). What is social marketing? www.thensmc.com/content/what-social-marketing-1

UNESCO-IHP, NARBO, WWAP. (2009). IWRM guidelines at River Basin level. http://www.unesco.org/water/

Voinov, A., & Bousquet, F. (2010). Modelling with stakeholders. *Environmental Modelling and Software, 25*(11), 1268–1281.

Index

A

Adaptive management, 12, 40, 43, 49, 88, 334, 410, 458, 459, 568, 578, 591, 594

Advective transport, 425, 426, 429

Aeration, 537

Agriculture, 12, 37, 41, 138, 469, 511, 567, 568, 570, 571, 574, 575, 585, 613

Algae eutrophication modelling, 438

Allocations, 30, 75, 83, 85, 97–99, 101, 102, 105–111, 113–115, 155, 166, 167, 169, 172, 175, 188–190, 197, 207, 209, 210, 305–307, 311–313, 315, 319, 329, 386, 387, 389, 458, 469, 505, 518, 583

Alternatives, 6, 34, 35, 38, 40, 45, 60, 69, 73, 76, 79, 81, 82, 88, 94, 96, 151, 163, 233, 270, 327, 346, 349, 366, 376–379, 394, 397, 398, 402, 414, 465, 490, 530, 596, 610

Anaerobic, 438, 441, 555, 557

Annual value, 96, 137, 313

Aquatic systems, 428

Aquifer, 2, 5, 6, 31, 38, 67, 71, 151, 258, 379, 568, 574

Artificial neural networks, 55, 63, 179–181, 183, 207

Autocorrelations, 251, 254, 256, 257, 260, 270, 299, 303

Autoregressive-moving average (ARMA) models, 269, 270

Average values, 215, 367, 425, 439

B

Benefit-cost analyses ratio, 165, 583

Benefits, 1, 16, 20, 24, 30, 33, 35, 37, 38, 41, 62, 75, 83, 85, 90, 94–96, 98, 99, 102, 107–109, 111, 114, 137, 155, 156, 158, 169, 171, 196, 210, 259, 288, 293–295, 307, 328, 373, 380, 383, 385–387, 393, 474, 493, 504, 505, 508, 512–514, 519, 520, 522, 558, 592

Bias, 183, 186, 211, 220, 222, 229, 256, 290, 297

Biochemical oxygen demand (BOD), 433, 436, 462, 537, 555

Boundary conditions, 104, 336, 338, 572, 573, 576, 585, 595, 607

Box-Jenkins models, 269

Budget, 58, 385, 512, 579, 580, 583, 587, 598

C

Calibration, 55, 80, 144, 145, 147, 179–181, 336–338, 340–342, 354, 360, 370, 425, 473, 536, 543, 559, 606

Capacity expansion, 75, 115, 116, 155, 174, 512, 514, 515

Capital, 24, 74, 95, 172, 216, 292, 387, 388, 559

Capital recovery factor, 96

Case studies, 2

Censored data, 245

Central limit theorem, 357

Chance constraints, 306, 327, 329

Climate change, 2, 527, 589, 590, 594, 595

Coasts beaches
    management, 2, 5, 40, 41, 43, 207, 425
    zones, 2, 40, 43, 580, 589

Coefficient of skewness, 220, 222, 223, 225, 235, 237–240, 244, 263, 290

Coefficient of variation, 220, 224, 225, 236, 240, 241, 248, 256, 290

Coliform bacteria, 537

Collective goods, 382

Compensation criterion, 382

Competitive conditions, 381

Concave functions, 98, 157, 160, 161

Concentration gradient, 426, 427

Confidence intervals, 231, 233, 277, 337

Constituents conservative
    first-order, 433

Constraint method, 395–397, 412, 415

Constraints, 71, 75, 79, 85, 102–105, 121, 129, 138, 142, 149, 151, 158, 163, 176, 211, 243, 307, 317, 327, 339, 378, 389, 402, 464, 477, 482–484, 486, 506, 508, 511, 515, 525, 559, 576, 586, 588, 606

Continuous variables, 339, 353

Convex functions, 101, 159–161

Correlation, 233, 234, 248, 251, 255, 260–264, 266–270, 289, 292, 294, 298, 299, 302, 303, 308, 342, 357–359, 555

Cost, 2, 10, 31, 34, 37, 41, 73, 76–79, 94–96, 116, 117, 119, 120, 146, 148–150, 153, 155, 156, 165, 166, 168, 169, 172–174, 177, 196, 211, 292, 293, 338,