Water Resources Systems Analysis

with Emphasis on Conflict Resolution

Mohammad Karamouz
Ferenc Szidarovszky
Banafsheh Zahraie

WATER RESOURCES SYSTEMS ANALYSIS

Mohammad Karamouz
Ferenc Szidarovszky
Banafsheh Zahraie

LEWIS PUBLISHERS

A CRC Press Company
Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Karamouz, Mohammad.
Water resources systems analysis / Mohammad Karamouz, Ferenc Szidarovszky, Banafsheh Zahraie.
p. cm.
Includes bibliographical references and index.
ISBN 1-56670-642-4 (alk. paper)
1. Water resources development. 2. System analysis. I. Szidarovszky, Ferenc. II. Zahraie, Banafsheh. III. Title.

TC409.K37 2003
333.91—dc21

2002043472

Visit the CRC Press Web site at www.crcpress.com

Lewis Publishers is an imprint of CRC Press LLC

No claim to original U.S. Government works
International Standard Book Number 1-56670-642-4
Library of Congress Card Number 2002043472
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Preface

In the past decades, the management of water resources has become more focused on mid- and long-term planning for water conservation and demand management, diversion and water transfer, and all phases of water resources development projects. It has become clear that structural solutions are not always the most efficient and economically sound alternatives. Nonstructural means of water resources management go beyond utilization of existing structures and the physical limits of water resources systems to include technical, social, political, and economic aspects for better resource management.

In this book, applications of a systems approach to water resources planning and management are presented, and potential conflicts between various sectors and disciplines affected by the development projects are discussed. Water resources systems analysis involves modeling of quantity and quality aspects of surface water and groundwater. Understanding the principles of simulation, optimization techniques, and multiple-criterion decision making as well as engineering economics and time series analysis are prerequisites for water resources systems analysis (see Chapters 2 to 5). The book is written for graduate courses in water resources in civil, environmental, and agricultural engineering studies. The objective of each chapter has been to incorporate the latest developments in surface and groundwater systems optimization and modeling. The modeling techniques are discussed with some field applications.

Chapter 1 presents basic definitions about hydrologic and environmental systems. This chapter also summarizes major conflict issues and the framework for a sustainable decision support system in water resources planning and management. Chapters 2 and 3 summarize the methods and concepts of decision making and elements of probabilistic and statistical methodology, as well as their use in modeling risk and uncertainty. Mathematical methods for conflict resolution are also explained in these chapters.

Chapter 4 presents the principles of engineering economics and their application to water resources planning and operation, as well as methods for incorporating the time/money relationship. Chapter 5 discusses basic principles of hydrologic time series modeling and types of statistical models related to water resources. Chapter 6 summarizes various elements of river basin modeling, including data and information processing, as well as objectives, constraints identification, and details of the modeling process. Management tools such as simulation and optimization models and a number of computer packages that have been used in water resources planning and management projects are also presented in this chapter.

Chapter 7 discusses the main characteristics of groundwater systems, groundwater flow equations, and groundwater modeling and provides applications of optimization models in groundwater management. Chapter 8 summarizes various phases

of river–reservoir systems modeling, including data and information processing, and deals with such issues as water supply, flood control, power generation, and instream flow requirements. Formulation of simulation and optimization models for water resources planning and management in river–reservoir systems are also presented in this chapter.

Chapter 9 discusses principles of water quality analysis, including major types of pollutants, water quality criteria, and water quality monitoring. Water quality management in rivers and reservoirs is discussed in the second part of this chapter, and groundwater pollutants and principles of groundwater quality management are also covered. Chapter 10 presents major concerns of water resources engineers regarding the design and operation of hydroelectric power plants; it also provides a framework for long- and short-term operation models. Coordination between hydro and thermal power plants is also discussed in this chapter. Chapter 11 summarizes various aspects of managing water demands for domestic, industrial, and agricultural purposes. Wastewater treatment and water reuse issues as they relate to water management are also discussed in this chapter. Chapter 12 defines climatic, hydrologic, and agricultural droughts, and the basic steps for analyzing drought for drought management studies are demonstrated.

It is our hope that this book can serve water resources communities around the world and add significant value to systems analysis techniques for water resources planning and management.

Mohammad Karamouz

Ferenc Szidarovszky

Banafsheh Zahraie

Acknowledgments

Many individuals have contributed to the preparation of the book. The initial impetus for the book was provided by Professors Mark H. Houck and Jacques W. Delleur through their class notes provided when they taught at Purdue University. We are indebted to Mr. Reza Kerachian, Ph.D. Candidate of Amirkabir University (Tehran Polytechnic), for his dedicated and considerable contribution to Chapters 6, 7, 8, and 9. The authors would also like to thank other doctorate students of Amirkabir University (Tehran Polytechnic); Shahab Araghy-Nejad for his contribution to Chapters 5 and 12; Ali Moridi for his contribution to Chapter 11; Koroosh Behzadian for his contribution to Chapter 10; and Arman Ganji of Shiraz University for his contribution to conflict resolution example problems. Special thanks to Neda Parvini for preparing figures and artwork for the book. The reviews and constructive comments of Professors Richard McCuen, Mark H. Houck, Lindell Ormsbee, Nathan Buras, G. V. Loganathan, and Dr. Charles Howard are greatly appreciated.

About the Authors

Mohammad Karamouz is Professor of Civil Engineering at Amirkabir University (Tehran Polytechnic) and a visiting professor in the Department of Civil Engineering/Hydrology and Water Resources at the University of Arizona, Tucson. He received a B.S. degree in civil engineering from Shiraz University, an M.S. in water resources and environmental engineering from George Washington University, and a Ph.D. in hydraulic and systems engineering from Purdue University. He is a Licensed Professional Engineer in the state of New York and a Fellow of American Society of Civil Engineers. Among many professional achievements, he was the dean of engineering at Pratt Institute and the founder of Arch Consulting Co. in New York City. His areas of research and teaching interests include integrated water resources planning and management, drought analysis and management, decision support systems, streamflow forecasting, reservoir operation, conjunctive use of surface and groundwater, and genetic algorithm modeling. He has more than 100 technical publications, books, and book chapters. Professor Karamouz serves nationally and internationally as a consultant to both private and governmental agencies.

Ferenc Szidarovszky is a graduate of the Eotvos University of Sciences, Budapest, Hungary. He has two Ph.D. degrees, one in mathematics and the other in economics. The Hungarian Academy of Sciences has awarded him with the degrees Candidate in Mathematics as well as Doctor of Engineering Sciences. After moving to the U.S., he became a professor in the Systems and Industrial Engineering Department at the University of Arizona, Tucson. His research interests include numerical analysis, systems theory, decision theory, including multi-criteria decision making, conflict resolution, and their applications in natural resources management and economics. He is the author of more than 200 research publications and authored or co-authored 8 books in Hungarian and 9 books in English.

Banafsheh Zahraie is an assistant professor at the Department of Civil Engineering at the University of Tehran. She received a B.S. in civil engineering, an M.S. in water resources engineering, and a Ph.D. in water resources management from Amirkabir University (Tehran Polytechnic). She was a research associate at the Department of Hydrology and Water Resources at the University of Arizona through a World Lab fellowship. She is an international expert in application of systems engineering in hydropower reservoirs and has developed many state of the art computer-modeling packages and decision-support systems for large-scale water resource systems. She is a member of American Society of Civil Engineers and American Geophysical Union. Her research interests include decision support systems, MCDM, stochastic hydrology, and systems analysis.

Contents

1 Introduction

1.1 INTRODUCTION

Increasing demand for water, higher standards of living, depletion of resources of acceptable quality, and excessive water pollution due to agricultural and industrial expansions have caused intense social and political predicaments. Since the beginning of the 19th century, the population of the world has tripled, nonrenewable energy consumption has increased by a factor of 30, and the industrial production has multiplied by 50 times. Although progress has improved the quality of life, it has caused significant environmental destruction in such a magnitude that could not be predicted.

A question that should be addressed is whether in the next decades development could be done in a way that is economically and ecologically sustainable. We cannot answer this question unless we have a vision of the future and our planning schemes are environmentally responsible and sensitive toward the major elements of our physical environment, namely air, water, and soil. Among these elements, water is of special importance. Excessive use and misuse of surface and groundwater and pollution of these vital resources by residential, agricultural, and industrial wastewater has threatened our well-being. Planning for sustainable development of water resources means water conservation, waste and leakage prevention, improved efficiency of water systems, improved water quality, water withdrawal and usage within the limits of the system, a level of water pollution within the carrying capacity of the streams, and water discharge from groundwater within the safe yield of the system. In other words, we are seeking a balance among our physical being, our ability to manage our resources, and the limitations imposed by the environment.

Water is a sustainable resource, and the need for integrated water resources management is on the agenda of every state. As one of the three vital elements of air, soil, and water, the role of water in the conservation of the environment is receiving increased global attention. Figure 1.1 illustrates the vulnerability of various nations to water scarcity by using a composite index based on available water resources and current use, population, reliability of water supply, and national income (World Resources, 1998).

In general, water resources development studies can be classified into planning, operation, and management. Long-term management of water resources systems requires considering the competing users of the system along with the future supply and demands. Then, by taking into account these considerations, modeling operation of the systems will be helpful to the decision makers in charge of allocating resources with higher efficiency. Systems need to be expanded because of the growing demands for resources and services; therefore, capacity expansion is another part of the long-term water resources development studies when existing

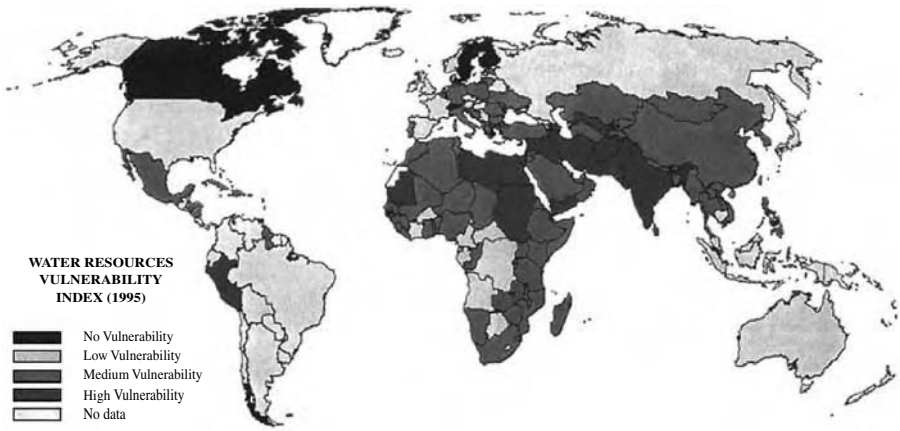


FIGURE 1.1 Vulnerability of nations to water scarcity. (From World Resources Institute, *World Resources: A Guide to the Global Environment*, a joint publication by The World Resources Institute, The United Nations Environment Program, The United Nations Development Program, and The World Bank, Oxford University Press, New York, 1998.)

facilities for storing, withdrawing, and transferring water to users are not adequate with an acceptable reliability.

Figure 1.2 shows different steps in the water resources planning process. As shown in this figure, the process can be classified into the following phases:

1. Problem definition and data collection and processing
2. Modeling
3. Decision making
4. Construction
5. Continuous monitoring of the system

In the first phase, the objectives and constraints are identified. Data collection and processing are important parts of this process and include:

- Climatic, hydrologic, hydraulic, and environmental characteristics of the study area affecting the supply and the demands
- Economic, institutional, and legal conditions affecting water allocation policies
- Structural and physical characteristics of the rivers and reservoirs and their appurtenant facilities that have impacts on the carrying capacity of the system

In the modeling phase, simulation and optimization models are used for finding possible alternatives. In the next phase, conflict resolution and multi-criterion decision analysis techniques are used for comparing different alternatives. As seen in Figure 1.2, the selection of final alternatives, in many cases, depends on social and

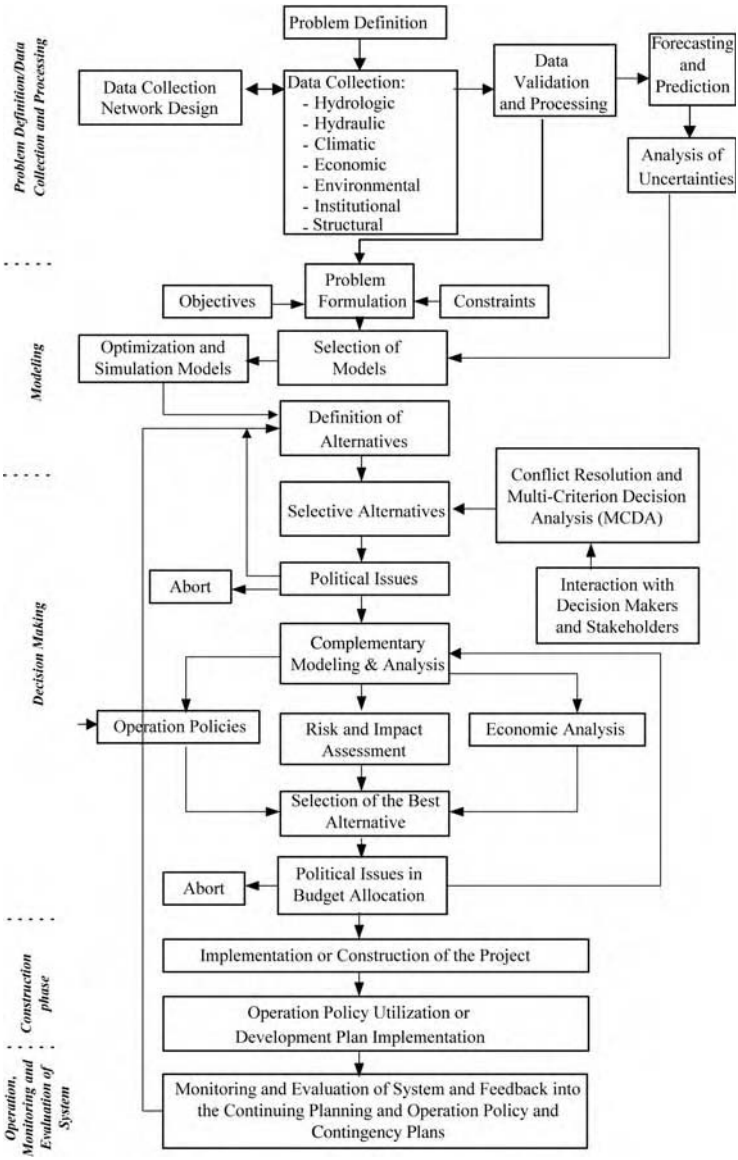


FIGURE 1.2 Process of water resources planning.

economic (nontechnical) issues related to institutional framework and distribution of financial resources in a region. Political issues may also cut off the process if the policy makers do not effectively participate in the planning process. In the last phase, monitoring, evaluation, and feedback to decision makers are carried out in order to modify the planning schemes and the operating policies.

TABLE 1.1
Water Reserves on the Earth

	Volume (10^{12} m^3)	Percentage of Global Reserves (%)	
		Total Water	Freshwater
World ocean	1,338,000	96.5	—
Groundwater	23,400	1.7	—
Freshwater	10,530	0.76	30.1
Soil moisture	16.5	0.001	0.05
Glaciers and permanent snow cover	24,064	1.74	68.7
Antarctic	21,600	1.56	61.7
Greenland	2340	0.17	6.68
Arctic islands	83.5	0.006	0.24
Mountainous regions	40.6	0.003	0.12
Ground ice/permafrost	300	0.022	0.86
Water reserves in lakes	176.4	0.013	—
Fresh	91	0.007	0.26
Saline	85.4	0.006	—
Swamp water	11.47	0.0008	0.03
River flows	2.12	0.0002	0.006
Biological water	1.12	0.0001	0.003
Atmospheric water	12.9	0.001	0.04
Total water reserves	1,385,984	100	—

Source: Shiklomanov, I., in *Water in Crisis*, P. Gleick, Ed., Oxford University Press, New York, 1993, chap. 2. With permission.

1.2 AVAILABILITY OF WATER ON EARTH AND WATER USE

As can be seen in Table 1.1, oceans, the greatest water reserve on Earth, represent 96.5% of the total reserve. An examination of the other reserves shows that shares of permanent snow covers and groundwater that are major sources of freshwater add up to 3.44% of the total water reserve. The spatial distribution of water availability in different regions can be a detriment to regional and local development plans.

The total world freshwater supply is estimated to be 41,022 billion m^3 (World Resources, 1998). Table 1.2 shows the freshwater supply of different continents and annual withdrawals for domestic, industrial, and agricultural water uses. As can be seen in this table, annual withdrawal is about 8% of the total freshwater supply, and the agriculture sector, with 69% usage, has the highest rate among the water uses.

1.3 WATER RESOURCES SYSTEMS: BASIC CONCEPTS

Each *water resources system* consists of different elements of two distinct environments: one is a *physical, chemical, and biological environment* and the second is a *cultural environment* with social, political, economic, and technological dimensions

TABLE 1.2
Total Water Supply in Different Continents

Continents	Annual Internal Water Resources[a] (10^{12} m^3)	Year of Data	Annual Withdrawal		Sector Withdrawal		
			Total (10^{12} m^3)	Water Resources (%)	Domestic (%)	Industrial (%)	Agricultural (%)
Africa	3.996	1995	0.15	4	7	5	88
Europe	6.235	1995	0.46	7	14	55	31
North America	5.309	1991	0.51	10	13	47	39
Central America	1.057	1987	0.10	9	6	8	86
South America	9.526	1995	0.10	1	18	23	59
Asia	13.206	1987	1.63	12	6	9	85
Oceania	1.614	1995	0.02	1	64	2	34
World	41.022	1987	3.240	8	8	23	69

[a] The average annual flow of rivers and recharge of groundwater generated from endogenous precipitation.

Source: World Resources Institute, *World Resources: A Guide to the Global Environment*, a joint publication by The World Resources Institute, The United Nations Environment Program, The United Nations Development Program, and The World Bank, Oxford University Press, New York, 1998.

(White et al., 1992). The physical environment includes interdependent water bodies and structures, each impacting the state and performance of the others. The cultural environment encompasses the various social constraints that are mainly focused on us and our interactions with physical environment. The physical and cultural environments are inseparable. In other words, natural resources and especially water resources systems cannot be modeled effectively without considering social constraints. Over the course of this book, we try to show how different elements of these two environments and their interactions are in conflict and how they can be modeled.

For modeling, one of the most common approaches is to divide the system into pieces (*disintegration phase*) to see what it is composed of and how it works. It is an example of the reductionist approach, which reduces complex phenomena to their component parts and examines them in detail. The problem with this approach becomes apparent when we try to put the system back together again. The process of turning our focus from analysis to synthesis can become more practical by considering the following items:

- The interactions among components and their role in the entire system should be defined before the system is broken into pieces. Knowledge about these interactions plays an important role in the disintegration phase and in the analysis of the entire system.
- The approaches that are used for modeling different components should be consistent.
- The terminologies and methodologies for defining and assessing the performance of each component should be comparable.

The second phase in systems analysis is to integrate different components of the system in order to assess the performance of the system as a whole. Integrated water resources planning and management focus not only on the performance of individual components but also on the performance of the entire system of components (Loucks, 1996).

1.4 SYSTEM DEFINITION AND PROPERTIES

Any structure or device, including different interactive components (real or abstract), that causes an output reference to a specific input in a given time can be called a system. The common characteristics of any system in general and water resources systems specifically can be summarized as follows:

- All systems have some structure and organization.
- Systems are all generalizations, abstractions, or idealizations of the real world with different levels of complexities.
- Functional and structural relationships exist between components of the system.
- All systems show some degree of integration.
- Input–output relations and the nature of them are important characteristics of systems.

Functional relationships reflect both physical and social functioning of the system. Physical functioning includes the flow and transfer of some material, while social functioning involves social, political, and economic preferences. Specifications of systems can be classified as follows:

- Inputs
- Governing physical laws
- Initial and boundary conditions
- Outputs

Inputs, outputs, and major characteristics of the systems are usually defined by variables and parameters. A *variable* is a system characteristic that assumes different values measured at different times. A *parameter* is also a system characteristic, but it normally does not change with time. Different values of variables can be viewed as possible *states of the nature* or as alternative futures. Variables or parameters can be classified as:

- *Lumped*, which do not change in space
- *Distributed*, which vary in one or more space dimensions

Memory, which is another important characteristic of a system, is the length of time in the past for which the input could have an impact on the output. A system may have different levels of memory:

- *Zero memory* — The state and output depend only on the input in the present time.
- *Finite memory* — The state, output, and behavior depend only on the history of the system for a specific time span (memory)
- *Infinite memory* — The state and output depend on the entire history of the system

Systems are generally classified as deterministic or stochastic. In *deterministic systems*, the same input gives the same output, whereas *stochastic systems* contain one or more elements for which the relationship between input and output is probabilistic rather than deterministic.

There are also other detailed classifications of systems. They are as follows:

- *Continuous systems* — The output is produced continuously.
- *Discrete systems* — The output changes after finite intervals of time.
- *Quantized systems* — The output values change only at certain discrete intervals of time and hold a constant value between these intervals.
- *Natural systems* — The inputs and outputs and other state variables are measurable and are not controlled.
- *Devised systems* — The input may be both controllable and measurable.
- *Simple systems* — No feedback mechanisms exist in these systems.
- *Complex systems* — Feedback is built into these systems.

- *Adaptive systems* — These systems learn from their past history to improve their performance.
- *Causal systems* — An output cannot occur earlier than the corresponding input (cause and effect).
- *Simulation systems* — These are realization systems and are similar to causal systems.
- *Stable systems* — If the input is bounded, the output is also bounded and vice versa.
- *Damped systems* — The output of the system dies out without ever crossing the time scale.

Most of the hydrologic systems are stable and causal systems and are heavily damped. Figure 1.3 shows examples of outputs of continuous, discrete, and quantized systems.

More details about continuous and discrete variables are presented in the following sections of this chapter. Considering the three basic components of a system (inputs, system operators, and outputs), as shown in Figure 1.4, systems analysts could face several different types of problems:

- *Design problems* — Inputs and system are known and the output must be quantified.
- *System identification problems* — Inputs and outputs are known and the system itself must be identified.
- *Detection problems* — System and outputs are known and inputs must be identified.
- *Synthetic problems (simulation)* — Inputs and outputs are known and the performance of models must be tested.

Hydrologists primarily deal with design and synthetic problems.

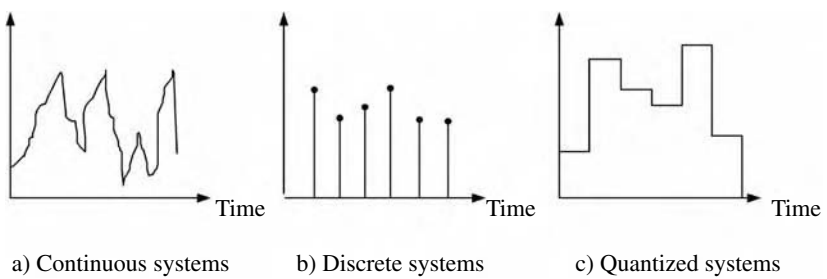


FIGURE 1.3 Outputs of continuous, discrete, and quantized systems.

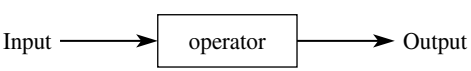


FIGURE 1.4 Schematic representation of systems operation. (From Chow, V. T., Maidment, D. R., and Mays, L. W., *Applied Hydrology*, McGraw-Hill, New York, 1988. With permission.)

1.5 HYDROLOGIC SYSTEMS AND MODELING APPROACHES

Complex systems can be decomposed into *subsystems,* each having an input–output linkage as a component. *Hydrologic systems* can be considered as a subsystem of water resources representing the physical functioning of that system in a region. A hydrologic system is defined as a structure or volume in space, surrounded by a boundary, that accepts water and other inputs, operates on them internally, and produces them as outputs (Chow et al., 1988).

Two approaches can be used for modeling hydrologic systems:

- Theoretical approach, which focuses on modeling the physics of the system
- Empirical approach, which focuses on using historical observations of different components of the hydrologic cycle

Utilizing the first approach results in more accurate models and outputs, thus improving our knowledge about different hydrologic phenomena; however, modelers are usually faced with a lack of accurate information about inputs of the system and a wide range of natural complexities.

Models are simplified representations of systems. Loucks et al. (1981) refer to model development as an art that requires judgment in abstracting from the real world of the components that are important to the decision-making process. Mathematical models, which are the main focus of this book, can be classified into the following categories:

- Empirical vs. theoretical
- Lumped vs. distributed
- Deterministic vs. stochastic
- Linear vs. nonlinear

Details of these models are explained in different chapters of this book. Before proceeding with the details of systems of analysis, the holistic view of water resources planning and management could only be meaningful if it is observed within the context of the hydrologic cycle, at least in a catchment scale.

1.5.1 Hydrologic Cycle

The *hydrologic cycle* is defined as "the pathway of water as it moves in its various phases through the atmosphere to the Earth, over and through the land, to the ocean, and back to the atmosphere" (National Research Council, 1991). The hydrologic cycle can be considered as a closed system for Earth, because the total amount of water in the cycle is constant.

The hydrologic cycle is generally described in terms of six major components and some subcomponents, which are shown in Figure 1.5:

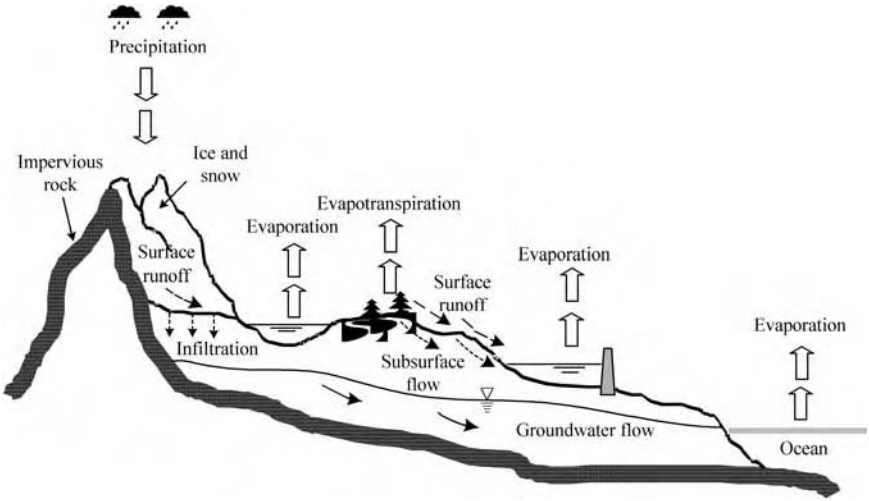


FIGURE 1.5 Schematic diagram of the hydrologic cycle.

- Precipitation (P)
  - Precipitation on the oceans (P_o)
  - Precipitation on ice (P_i)
  - Precipitation on land (P_l)
- Infiltration (I)
- Evaporation (E)
  - Evaporation from oceans (E_o)
  - Evaporation from ice (E_i)
  - Evapotranspiration from plants (T_s)
  - Evaporation from soil moisture (E_g)
- Surface runoff (R)
- Goundwater appearing as surface runoff (R_g)
- Subsurface flow (R_s)
- Groundwater flow (G)

Water budget is an accounting of the inflow, outflow, and storage of water in a specific hydrologic system. Figure 1.6 is a schematic diagram of a system representing the hydrologic cycle of a watershed. As can be seen in this figure, R_{in}, G_{in}, R_{out}, and G_{out} are surface and groundwater inflows and outflows, respectively. Surface inflow in this figure represents an inter-basin water transfer. R_g and E_g are groundwater appearing as surface runoff and evaporation from soil moisture, respectively.

The surface water budget (ΔS_s) and groundwater budget (ΔS_g) can be estimated as follows:

$$\Delta S_s = P_l + R_{in} - R_{out} + R_g - T_s - I \tag{1.1}$$

$$\Delta S_g = G_{in} - G_{out} - R_g - E_g + I \tag{1.2}$$

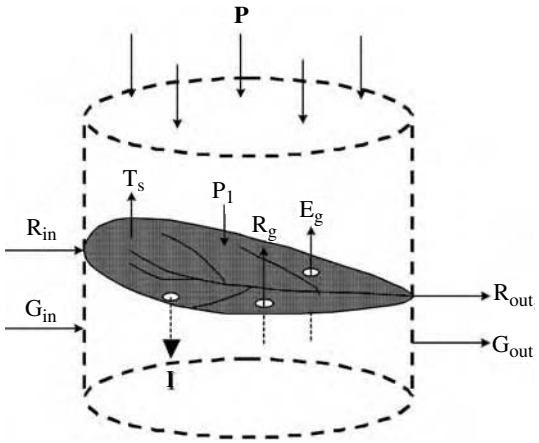


FIGURE 1.6 Schematic diagram of a system representing the hydrologic cycle of a watershed.

1.5.2 Hydrologic Variables and Parameters

All natural physical processes are subject to variability. For example, rainfall intensity, flood magnitude, or low flows in droughts have wide variations. To study these variations and incorporate them in the planning and operation of water resources, engineers gather and investigate *samples* of data.

The hydrologic cycle is composed of various phenomena, such as precipitation, runoff, infiltration, evaporation, evapotranspiration, and abstraction. Different *characteristic variables*, which can simply be called *hydrologic variables*, have been defined to describe each of these phenomena. Depth or intensity of rainfall in different time steps of a rainstorm, monthly inflow discharge to a reservoir, or daily evaporation are some examples of hydrologic variables (Shahin et al., 1993). A dataset consists of a number of measurements of a phenomenon, and the quantities measured are variables.

A *continuous variable* can have any value on a continuous domain; examples include volume of water flowing in a river or the amount of daily evaporation measured in a climatic station. A *discrete variable* represents an interval or the number of occurrences within each interval of time and space; the number of rainy days in a certain period of time (e.g., a year) is an example of discrete hydrologic variables.

The hydrologic variables can also be classified as qualitative or quantitative. A *qualitative variable* can be expressed as a real number in a sensible way; type of soil is an example of a qualitative hydrologic variable. A *quantitative* variable can be measured as a real number; the number of rainy days in a year and rainfall intensity in a day are examples of discrete and continuous quantitative hydrologic variables, respectively.

Hydrologic variables usually vary in time and space. A *time series* is a sequence of values arranged in order of their occurrence in time. Conflicts can arise when trying to analyze the impact of various hydrologic variables on each other and model them during the decision-making process.

1.6 CONFLICT ISSUES IN WATER RESOURCES PLANNING AND MANAGEMENT

Water resources systems usually have multiple objectives, many of which are usually in conflict. For example, a major conflict issue in the operation of river–reservoir systems arises when the reservoir is not capable of supplying all the demands; supplying instream requirements is in conflict with water conservation and supply objectives, and monthly variations of power loads (energy demand) and water demands do not follow the same pattern of decisions. For example, the operation of groundwater resources can be a multifaceted problem with many economic, hydrologic, hydraulic, water quality, or environmental objectives and constraints that must be satisfied. A common conflict issue in groundwater systems planning and operation arises when the aquifer supplies water to different demand points for different purposes and constraints on groundwater table fluctuations and quality aspects must be maintained. Let's consider the following example:

- Several demand points exist and groundwater that is supplied to one of these demand points cannot be used by the others; therefore, a major conflict issue may occur.
- Excessive discharge of aquifer can cause problems, such as settlement of ground surfaces and buildings.
- Recharge of aquifer with polluted water such as infiltration of agricultural wastewater, disposal of sewage or sludge from water treatment plants, and sanitary landfills can produce environmental problems especially for aquifers intended for water supply purposes and causes some irreversible conflicts.

The optimal operating policies for operation of groundwater resources can be developed using hydraulic or water quality response equations, well characteristics, and hydraulic gradient and water demand requirements, as well as conflict resolution models that consider the supply, demand, and physical characteristics of the systems. Besides conflicts in the objectives of water users, other important issues must be considered from an institutional point of view. Even though water resources planning for surface and groundwater operation might be done by a specific agency, such as the department of water supply, the water allocation schemes defined and imposed by this particular agency could be highly affected by the political and institutional strength of other agencies, such as the Department of Agriculture, various industries, and the Department of Environmental Protection. On the other hand, these and many other entities are affected by water policies because they:

- Use water resources for producing specific goods and products.
- Use water resources for providing some services.
- Are responsible for public health and environmental protection.

Table 1.3 presents an example of how the quantitative aspects of conflict issues in water supply between different agencies can be quantified. This example is based

on organizational setups in some developing countries. The qualitative aspects of conflict issues are shown in Table 1.4. As can be seen in these tables, the various agencies that directly use water or might be affected indirectly by water supply policies are in conflict in varying degrees.

The priorities and favorable ranges of water supply for each of these agencies should be considered in formulating a conflict resolution scheme; therefore, the first step in conflict resolution studies is to recognize all of the conflict issues and the responsible agencies. Each of these departments has its own set of priorities for allocating water to different demands. When formulating the conflict-resolution algorithm, methods such as Nash bargaining theory (NBT) and multiple-criteria decision making (MCDM) can be used (see Chapter 2).

These methods can provide the water resources systems analyst with a better understanding of the mechanisms of cooperation, the roots of and the reasons for conflicts, and the resolutions required in various political and social settings. They can also promote a better understanding of the issues and help the stakeholders to comprehend better with the potential consequences of preferences and choices. Throughout this book, conflict issues as they relate to the topics of particular chapters are discussed, and the algorithms for resolving conflicts based on qualitative and quantitative measures are presented.

1.7 DECISION TOOLS IN WATER RESOURCES MANAGEMENT

In the past, different tools have been used for modeling water resources systems, including simulation and optimization models. Yeh (1985) presented a state-of-the-art review of reservoir management and operation models in general. Yakowitz (1982) reviewed the application of dynamic programming in water resources planning and management. These efforts have been mainly focused on development of efficient tools for decision making in the area of water resources development and management. The continuing evolution of information technology, including hardware and software developments, has created an environment for better utilization of these tools.

The application of a systems approach to water resources planning and management has been established as a progressive dimension in the field of water resources engineering. Two approaches have been significantly enhanced in recent years. The first one is focused on the water-related data availability of variables in time and space domain affecting the uncertainty of water resources decision making. The second approach focuses on the complexity of the water resources domain and the complexity of modeling tools in an environment characterized by rapid technological development (Simonovic, 2000).

The decision support system (DSS) is a powerful tool for application of a systems approach in real-world water resources planning and management. The basis of DSS development was developed in the 1960s and 1970s in the field of management information system (MIS). While MIS relies on working with databases, DSS combines software and hardware facilities designed by systems engineers to be used to

TABLE 1.3
Conflict among Agencies over the Quantitative Aspects of Water Supply

Departments and Agencies	Water Supply Quantitative Aspects (water demands)				
	Domestic	Agriculture	Health	Industry	Environment
Department of Energy/Water Supply (A)	E (4), B (3), C (4), M (4)	B (4), C (4), E (2)	E (4), M (3)	B (4), M (3)	B (3), C (3), E (2), M (4)
Industrial and mining sector (B)	E (4), A (3), C (3), M (3)	A (4)	E (3)	A (4), C (4), E (3), M (3), W (1)	M (4), A (3)
Department of Agriculture (C)	A (4), E (4), B (3), A (4), M (4)	A (4), M (3)	M (3), E (3)	B (4)	A (3), E (2), M (4)
Health Department (E)	B (3), C (4), A (4), M (4)	A (2)	A (4), B (3), C (3), M (3)	B (3)	A (2), C (2)
Department of Environmental Protection (M)	A (4), B (3), E (4), C (4)	C (3)	A (3), C (3)	A (3), B (3)	A (4), B (4), C (4)

Note: The degree of conflict is indicated in parentheses; $i = 5$ is the most severe, and $i = 1$ is the least severe.

TABLE 1.4
Conflict among Agencies over the Qualitative Aspects of Water Supply

	Water Supply Qualitative Aspects					
Departments and Agencies	**Domestic**	**Agriculture**	**Health**	**Industry**	**Environment**	**Fishery and Domesticated Animal**
Department of Energy/Water Supply (A)	E (4), M (3)	C (4), E (2), M (4)	E (4), M (3)	B (4), M (4), E (2)	C (3), E (3), M (4)	E (4), M (4)
Industrial and mining sector (B)	E (4), M (3)	C (3), E (3), M (4)	E (5), M (5)	A (4), C (5)	E (3)	E (4), M (4)
Department of Agriculture (C)	M (2), E (4)	A (4), B (3), M (2)	E (4)	B (5), E (3), M (4)	A (3), M (4)	E (3), M (4)
Health Department (E)	A (4), B (4), C (4)	A (2), B (3), M (3)	A (4), B (5), C (4)	C (3), A (2)	A (3), B (3)	B (4), A (4), C (3)
Department of Environmental Protection (M)	A (3), B (3), C (2)	C (2), E (3), B (4), A (4)	A (3), B (5)	A (4), C (4)	A (4), B (4), C (4)	A (4), B (4), C (4)

Note: The degree of conflict is indicated in parentheses; $i = 5$ is the most severe, and $i = 1$ is the least severe.

make better decisions. The concept of DSSs developed as a result of the intersection of two trends (Stohr and Konsynski, 1992). The first one is the growing belief that existing information systems, despite their success in automation of operating tasks in organizational setups, have failed to assist the decision makers with many higher level tasks. The second trend is continuous improvement in computing hardware and software that has made it possible to place meaningful computing power into the development of databanks and complex heuristics. Several attempts have been made to develop a more flexible framework for these systems. The spatial decision support system (SDSS), adaptive decision support system (ADSS), and intelligent decision support system (IDSS) are some new DSS developments.

Figure 1.7 shows an example of a DSS framework for sustainable water resources planning and management. Access or Oracle DBMS, Visual Basic, HTML, and/or ASP (Java) can be used to define interfaces and develop a decision support system with all elements of water resources including: supply, demand, institutions, knowledge base, and conflict assessment. This system consists of six modules as follow:

- Database management
- Streamflow modeling and forecasting
- Demand management (modeling)
- Operating models
- Drought management
- Real-time operation

As can be seen in Figure 1.7, the various types of data and information — including climatic, hydrologic, hydraulic, and physical characteristics of a water resources system, as well as assumptions and the results of planning and operation alternatives — are managed in the database management module. The streamflow modeling and forecasting module considers statistical and conceptual forecast models, as well as heuristics for forecast modification, by incorporating long-range climate signals. Built-in evaluation of hydroclimatic data for improved water supply forecasting, implementation of the probabilistic forecasts models, and correlation schemes of weather and streamflow data are the major capabilities of this module. In the demand management module the emphasis is on a reduction in the quantity of water that is unaccounted for, optimal crop mixture, irrigation management policies, and optimal policies for conjunctive use of surface and groundwater resources. Using these two modules, the decision maker is able to develop forecast models for water supply and demand and utilize them in real-time operation of water resources systems.

In the operating models module, optimization and simulation models are used to develop long-term and mid-term planning and operation management policies. The drought management module is designed in order to develop water resources management policies during droughts. Because of the occurrence of severe droughts in many countries in recent years, drought monitoring and management have been given more attention by investigators. In the real-time operation module, optimal policies and heuristics developed in different modules of the system will be applied in the real-time operation of water resources systems. Through the use of these two modules, the user will be able to define:

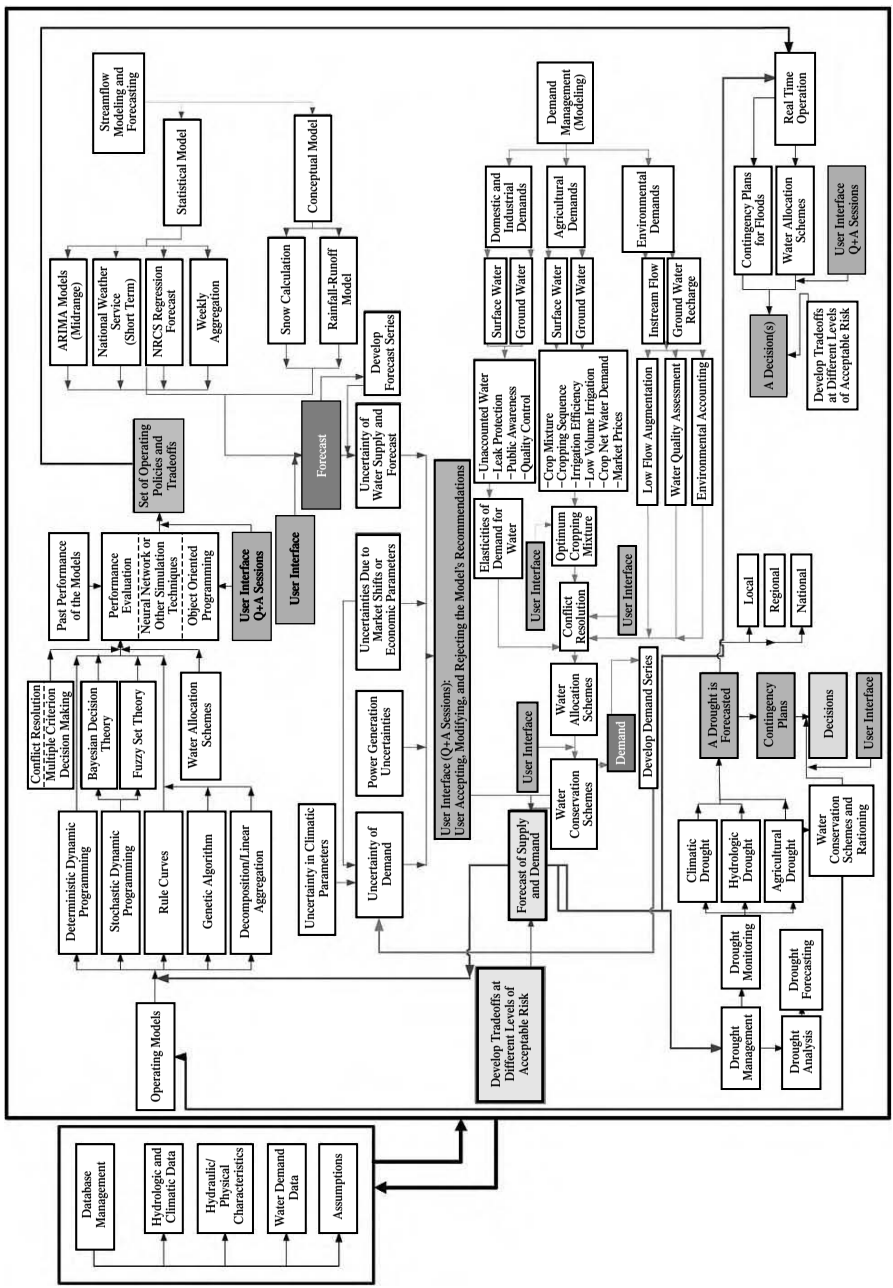


FIGURE 1.7 Framework of a decision support system (DSS) for sustainable water resources planning and management.

- Drought indicators
- Contingency plans for droughts and floods
- Water allocation schemes accepted by conflicting groups of water users
- Tradeoffs at different levels of acceptable risk

The DSS should incorporate recent advances in optimization under uncertainty and simulation and be capable of:

- Evaluation of alternatives
- Showing the impact of policies when different interest groups are searching for compromise solutions
- Incorporating uncertainty of supply and demand, uncertainty due to market shifts, and uncertainty due to inability of models to perform
- Incorporating parameter uncertainty and uncertainty due to vagueness of situations and/or boundaries
- Forecasting water supply and demand
- Surveying users' perceptions
- Supporting a variety of cognitive skills and styles and the level of knowledge of different users
- Assisting users to develop and use their own cognitive skills
- Developing schemes for changing public attitudes for sustainable water management
- Identifying new and improved scientific information for changing the market for water as an economic commodity
- Defining yardsticks to measure sustainability of the system with respect to the decisions

The system should be sufficiently flexible to allow the use and adaptation by users with various levels of experimental knowledge. Algorithms, heuristics, optimization, and simulation models in the DSS should be utilized in such a way that they examine the impact of various policies based on four different decision-making platforms:

- Decisions under certainty
- Decisions under risk
- Decisions under uncertainty
- Decisions under conflict

The DSS should also be empowered with a variety of techniques to identify the most effective methodologies for continuous feedback and interactive stakeholder input into the knowledge base. Different mathematical and computational tools should be considered in the various modules of the system, such as:

- Use of Bayesian decision theory to incorporate new data into the probabilistic decision process
- Use of fuzzy set theory to assign membership functions based on the opinion of the stakeholder, the analyst, and the decision maker about a given situation

- Utilizing genetic algorithms and other optimization techniques to search through alternatives/operating policies
- The use of neural networks and other simulation techniques to test the performance of different alternatives or scenarios and to develop forecast schemes
- The use of scenarios, object-oriented programming, multicriterion decision making, and simulation to test the ability of a decision-making system based on how it reacts to a given emergency situation such as a drought or flood

REFERENCES

Chow, V. T., Maidment, D. R., and Mays, L. W., *Applied Hydrology*, McGraw-Hill, New York, 1988.

Kottegoda, N. T. and Rosso, R., *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, McGraw-Hill, New York, 1997.

Loucks, D. P., Surface water resources systems, in *Water Resources Handbook*, Mays, L. W., Ed., McGraw-Hill, New York, 1996, chap. 15.

Loucks, D. P., Stedinger, J. R., and Haith, D. A., *Water Resources Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1981.

National Research Council, *Opportunities in the Hydrologic Sciences*, National Academy Press, Washington, D.C., 1991.

Shahin, M., Van Oorschot, H. J. L., and De Lange, S. J. D., *Statistical Analysis in Water Resources Engineering*, Applied Hydrology Monographs, Balkema, Rotterdam, 1993.

Shiklomanov, I., World fresh water resources, in *Water in Crisis*, Gleick, P., Ed., Oxford University Press, New York, 1993, chap. 2.

Simonovic, S. P., Tools for water management: one view of the future, *Water Int.*, 25(1), 76–78, 2000.

Stohr, E. A. and Konsynski, B. R., *Information Systems and Decision Processes*, IEEE Computer Society Press, Los Alamitos, CA, 1992.

White, I. D., Mottershead, D. N., and Harrison, S. J. (1992), *Environmental Systems: An Introductory Text*, Second ed., Chapman & Hall, New York.

World Resources Institute, *World Resources: A Guide to the Global Environment*, a joint publication by The World Resources Institute, The United Nations Environment Program, The United Nations Development Program, and The World Bank, Oxford University Press, New York, 1998.

Yakowitz, S., Dynamic programming applications in water resources, *Water Resources Res.*, 18(4), 673–696, 1982.

Yeh, W., Reservoir management and operation models: a state-of-the-art review, *Water Resources Res.*, 21(12), 1797–1818, 1985.

2 Decision Making: Optimization and Conflict Resolution

In our private and professional lives, decisions must often be made, and decision-making tools are commonly used in the quantitative sciences. This chapter gives an introduction to the major methods and concepts of this very important scientific field.

2.1 INTRODUCTION

Decision making is the science of choice. Selecting the best technology for a particular application, developing flood protection alternatives, and optimizing the operation of a reservoir are all problems of choice. The first element of any decision-making problem is the *decision maker.* For a particular problem, we might have a single person who is responsible for deciding what to do, or several people or organizations may be involved in the decision-making process. In the first case, we have only one decision maker; in the second case, multiple decision makers. When more than one decision maker is present, then they might have different priorities, objectives, and desires, thus no decision outcome is likely to satisfy every decision maker. In such cases, a collective decision has to be made, and the outcome depends on how the different decision makers take the interests of others into account. In other words, the outcome depends on their willingness to cooperate with each other.

The options from which the selection is made are called *alternatives* and the set of all possible alternatives is called the *decision space.* In many cases, the decision space has a finite number of elements. For example, selecting a technology from four possibilities results in a decision space with four elements. In many other cases, the decision alternatives are characterized by decision variables that represent certain values about which we are making a decision. For example, reservoir capacity can be any real value between the smallest and largest possibility. If the decision space is finite, then the decision-making problem is *discrete*, and if the decision alternatives are characterized by continuous variables then the problem is *continuous.*

All decision-making problems are faced with limitations resulting from financial considerations and technological, social, and environmental restrictions, among others. These limitations are called the *constraints* of the problem. The alternatives satisfying all constraints are *feasible*, and the set of all feasible alternatives is the *feasible decision space*. If the decision space is finite, then the construction of the feasible decision space is very simple. We have to check the feasibility of each

alternative by determining whether or not it satisfies all restrictions. If the alternatives are characterized by decision variables, the constraints are usually presented as certain equations or inequalities containing the decision variables.

Because we have to make a choice from a given set of feasible alternatives we need to measure how good those alternatives are. The goodness of any alternative can be characterized by its *attributes*. Most of these attributes can be described verbally, such as how clean the water is or how well it satisfies customers. In mathematical modeling, we need measurable quantities describing the goodness of each alternative with respect to the attributes. Such measures are the *criteria*. For example, water quality can be measured by pollutant concentrations, consumer satisfaction by supplies, and so on. In all decision problems, we want to accomplish or avoid certain things. To what degree we accomplish our goals and avoid unfavorable consequences should be among the criteria. In classical optimization models we usually have only one criterion to optimize; however, in most decision-making problems we are faced with several criteria that frequently can conflict with each other. We can assume that all criteria are maximized, because otherwise the criterion can be multiplied by –1. For example, cost and quality in any problem are conflicting criteria, as better quality requires higher cost. When only one criterion is involved, the problem is a *single-criterion* optimization problem. In the presence of more than one criteria, we have a *multiple-criteria* decision problem. In the case of multiple decision makers, we might consider the problem as multiple-criteria decision making, where the criteria of all the decision makers are considered to be the criteria of the problem.

2.2 SINGLE-CRITERION OPTIMIZATION

In this section the fundamentals of single-criterion decision-making are briefly outlined. For mathematical simplicity, let us introduce the following notations. Let X denote the feasible decision space, and let $x \in X$ be any feasible alternative. If X is a finite set, then the value of X can be any one of the number 1, 2, …, N, when N is the number of the elements of X. If X is a continuous set characterized by continuous decision variables, then $\underline{x}$ is a vector, $\underline{x} = (x_1, \ldots, x_n)$, where n is the number of the decision variables and $x_1, x_2, \ldots, x_n$ are the decision variables. If f denotes the single criterion, then we assume that f is defined on X and, for all $x \in X$, $f(x)$ is a real number. This condition is usually defined as $f : X \rightarrow R$. Because f is real valued, all possible values of $f(x)$ can be represented on the real line, and the set:

$$H = \left\{ f(x) | x \in X \right\}$$

is called the *criterion space*. Notice that the elements of X show the decision choices, and the elements of H give the consequences of the choices in terms of the values of the evaluation criterion. We can simply say that X shows what we can do and H shows what we can get.

If X is finite and the number of its elements is small, then by comparing the $f(x)$ values the best alternative can be easily identified. If the number of elements of X is large or even infinity but X is a discrete set (for example, when it is defined by decision variables with integer values), then the methods of discrete and combinatorial optimization are used. This methodology can be found in any textbook of advanced optimization techniques.

The most simple single-criterion optimization problem with continuous variables is *linear programming*. In order to give a general formulation, let n be the number of the decision variables and $x_1, x_2, \ldots, x_n$ be the decision variables. It is assumed that the criterion and all constraints are linear. The criterion is sometimes also called the *objective function*, and it is assumed to have the form:

$$f(x) = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n = \underline{c}^T \underline{x} \tag{2.1}$$

where $c_1, c_2, \ldots, c_n$ are real numbers, $\underline{c} = (c_1, \ldots, c_n)$, and

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \tag{2.2}$$

It is also assumed that all constraints are linear and all decision variables are nonnegative. If any of the decision variables have only negative values, then by introducing the new variable:

$$x_i^- = -x_i \tag{2.3}$$

it becomes nonnegative. If a decision variable can be both positive and negative, then we can introduce two new variables:

$$x_i^+ = \begin{cases} x_i & \text{if } x_i \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

and

$$x_i^- = \begin{cases} -x_i & \text{if } x_i < 0 \\ 0 & \text{otherwise,} \end{cases} \tag{2.5}$$

and then both of them are nonnegative, and

$$x_i = x_i^+ - x_i^- \tag{2.6}$$

Linear constraints are equalities or inequalities. Equality constraints have the general form:

$$a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n = b_i \tag{2.7}$$

and inequality constraints can be written as:

$$a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n \underset{\leq}{\overset{\geq}{or}} b_i \tag{2.8}$$

Next we show that all constraints can be rewritten as less than or equal to (≤) types of conditions. If it is a ≦-type constraint, then we have nothing to do. Constraints with ≥ inequalities can be multiplied by –1, and then the ≥ inequality changes to ≤. Equality constraint (2.7) is equivalent to two inequality constraints:

$$a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n \leq b_i$$

$$a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n \geq b_i$$

where the second relation can be multiplied by –1 to obtain:

$$-a_{i1}x_1 - a_{i2}x_2 - .... - a_{in}x_n \leq -b_i$$

By repeating the above procedure for all variables and constraints, if necessary, the resulting problem will have only nonnegative variables and ≤ type of inequalities in the constraints. Hence, any linear programming problem can be rewritten in this way to its *primal form*:

$$\begin{aligned}
&\text{Maximize} && c_1x_1 + c_2x_2 + \cdots + c_nx_n \\
&\text{Subject to} && x_1, x_2, \cdots, x_n \geq 0 \\
& && a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \leq b_1 \\
& && a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \leq b_2 \\
& && \qquad\vdots \\
& && a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \leq b_m
\end{aligned} \tag{2.9}$$

Note that when the objective function represents cost or contained pollutant amount, for example, smaller objective values represent better choices. In such cases, the objective function is minimized. In the above formulation, we assume maximization, as minimization problems can be rewritten into maximization problems by multiplying the objective function by –1.

In the very simple case of two decision variables, Eq. (2.9) can be solved by the graphical approach. As an illustration, consider the following example.

Example 2.1

The combination of three technologies is used to remove a certain pollutant from wastewater. The three technologies remove 1, 2, and 3 g/m^3 of the pollutants, respectively. The third technology variant seems to be the best, but it cannot be applied to more than 50% of the wastewater being treated. The costs of applying the technology variants are \$5, \$3, and \$2 per cubic meter. If 1000 m^3 must be treated in a day, and at least 1.5 g/m^3 of pollutant has to be removed, then a simple linear programming formulation can be used to model this optimization problem. Let x_1 and x_2 be the amount of wastewater (in m^3) for which technologies 1 and 2 are used. Then, $1000 - x_1 - x_2$ m^3 is the amount to which technology 3 is applied. Clearly,

$$x_1, x_2 \geq 0 \tag{2.10}$$

$$x_1 + x_2 \leq 1000 \tag{2.11}$$

The condition that the third technology cannot be used in more than 50% of the treated wastewater requires that

$$1000 - x_1 - x_2 \leq 500$$

that is,

$$x_1 + x_2 \geq 500 \tag{2.12}$$

The total removed pollutant amount (in grams) is:

$$x_1 + 2x_2 + 3(1000 - x_1 - x_2)$$

which has to be at least 1.5 g/m^3 of the total 1000 m^3 being treated, so:

$$x_1 + 2x_2 + 3(1000 - x_1 - x_2) \geq 1500$$

This inequality can be rewritten as:

$$2x_1 + x_2 \leq 1500. \tag{2.13}$$

The total cost is given as

$$5x_1 + 3x_2 + 2(1000 - x_1 - x_2) = 3x_1 + x_2 + 2000$$

Because 2000 is a constant term, it is sufficient to minimize $3x_1 + x_2$ or maximize $-3x_1 - x_2$. Constraint (2.12) is a ≥ type of relation which is multiplied by –1 in order to reduce it to a ≤ type of constraint. After this modification, we obtain the following primal-form problem:

$$\begin{aligned} \text{Maximize} \quad & f = -3x_1 - x_2 \\ \text{Subject to} \quad & x_1, x_2 \geq 0 \\ & x_1 + x_2 \leq 1000 \\ & -x_1 - x_2 \leq -500 \\ & 2x_1 + x_2 \leq 1500 \end{aligned} \tag{2.14}$$

Each constraint can be represented in the two-dimensional space as a half-plane, the boundary of which is determined by the straight line representing the constraint as an equality. For example, the equality $x_1 + x_2 = 1000$ is a straight line, with the x_1 intercept = 1000 and the x_2 intercept = 1000. Because the origin, the (0,0) point, satisfies the inequality, the half-plane that contains the origin represents the points that satisfy this inequality constraint. The intercept of the half-planes of all constraints gives the feasible decision space, as is shown in Figure 2.1. Notice that this set has five vertices: (500,0), (750,0), (500,500), (0,1000), and (0,500) and the set is the convex hull, the smallest convex set containing the vertices.

From the theory of linear programming, we know that if the optimal solution is unique, then it is a vertex, and in the case of multiple optimal solutions a vertex always exists among the optimal solutions. Therefore, it is sufficient to find the best vertex that gives the highest objective function value. Simple calculation shows that:

$$f(500,0) = -1500$$

$$f(750,0) = -2250$$

$$f(500,500) = -2000$$

$$f(0,1000) = -1000$$

$$f(0,500) = -500$$

and because –500 is the largest value, the optimal decision is $x_1 = 0$, $x_2 = 500$, and $1000 - x_1 - x_2 = 500$. The minimal cost is, therefore, $3x_1 + x_2 + 2000 = \$2500$.

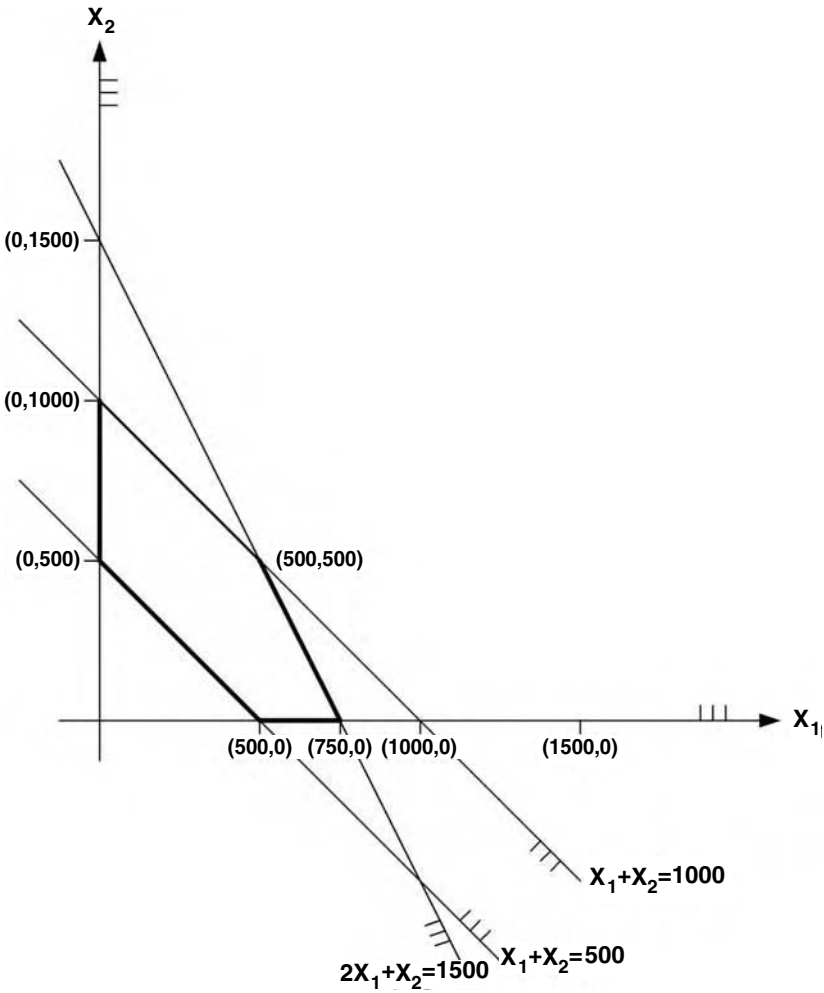


FIGURE 2.1 Illustration of the feasible decision space.

In the case of more than two decision variables, the graphical representation of the feasible decision space is impossible. A computational method known as the *simplex method* performs a systematic search on the vertices of the feasible decision space until the optimal vertex is found.

The concept of the simplex method can be summarized as follows: Any vertex of the feasible decision space corresponds to a basis of the column space of the coefficient matrix of the linear constraints. The variables associated with the columns of the basis are called *basic variables*. It can be shown that moving from one vertex to any of its neighbors is equivalent to exchanging one of the basic variables with a non-basic variable. In the simplex method, it is always guaranteed that in each step the objective function increases, which can be ensured by the following

conditions. Let a_{ij} be the (i,j) element of the coefficient matrix; b_i, the ith right-hand side number; and c_j, the coefficient of x_j in the objective function. Then, x_i is replaced by x_j if $c_j > 0$, $a_{ij} > 0$, and

$$\frac{b_i}{a_{ij}} = \min_l \left\{ \frac{b_l}{a_{lj}} \right\}$$

An initial basic can be found by introducing artificial variables to all ≥ and = type of constraints. For ≤ type of constraints, the slack variables play the same roles. In the first phase of the simplex method, all artificial variables are removed from the basis by successive exchanges, after which we have an initial feasible basic solution that is then successively improved in the second phase of the simplex method until optimum is reached. Many professional software packages implement the simplex method, so it is often used in practical applications.

In solving large-scale linear programming problems, we often encounter specially structured coefficient matrices. The most common structure is as follows:

$$\begin{pmatrix} A_{0,1} & A_{0,2} & \cdots & A_{0,k-1} & A_{0,k} \\ A_{1,1} & & & & \\ & A_{2,2} & & & \\ & & \ddots & A_{k-1,k-1} & \\ & & & & A_{k,k} \end{pmatrix}$$

where $A_{i,j}\ (0 \le i \le k, 1 \le j \le k)$ are given matrices. When applying the simplex method in such cases, we can take advantage of the special structure, and in each simplex step (moving to a neighboring vertex) we are able to preserve this special structure.

Another special method is the *decomposition* technique, where the optimal solution of the large problem can be formulated by using a sequence of the solutions of much smaller problems, the sizes of which are determined by the sizes of the blocks $A_{i,i}$ (i = 1, 2, …, k). The most popular decomposition technique is the Dantzig–Wolf method, a discussion of which can be found in almost all texts of optimization. The classical text of Dantzig (1963) is a good source for further reading.

If the objective function and one or more constraints are nonlinear, then the problem becomes one of *nonlinear programming*. In the case of only two variables, they can be solved by the graphical approach; however, the feasible decision space may not have vertices, and even if it does the optimal solution might not be a vertex. In such cases, the curves of the objective function with different values have to be compared. This procedure is shown next.

Example 2.2

For illustrative purposes, consider the previous example but assume that instead of minimizing the linear function $3x_1 + x_2$, a nonlinear function:

$$f(\underline{x}) = x_1^2 + x_2^2$$

is minimized. In order to find the feasible solution that minimizes this objective function, we first illustrate the solutions, which give certain special objective function values. The points with zero objective function values satisfy the equation:

$$f(\underline{x}) = x_1^2 + x_2^2 = 0$$

when the only solution is $x_1 = x_2 = 0$, or the origin. Points with an objective function value of 10,000 satisfy the equation:

$$f(\underline{x}) = x_1^2 + x_2^2 = 10,000$$

which is a circle with the origin at its center and a radius of 100. This circle has no feasible point, so 10,000 is a too small value for the objective function. By increasing the value of $f(\underline{x})$, we always get a similar circle with the same center, but with increasing radius (see Figure 2.2). By increasing the radius of the circle, we see that the smallest feasible radius occurs when the circle and the line $x_1 + x_2 = 500$ just touch each other. The tangent point is (250,250), which is therefore the optimal solution of the problem. In more complicated and higher dimensional cases, computer software must be used, and several professional packages are available to perform this task.

The most popular nonlinear optimization method is *linearization*, where all nonlinear constraints as well as the objective function are linearized if necessary. The resulting linear programming problem is then solved by the simplex method. Let $g(x_1, \ldots, x_n)$ be the left-hand side of a constraint or the objective function. Furthermore, let $(x_1^0, \ldots, x_n^0)$ be a feasible solution. Then, the linear approximation of g is given by the linear Taylor's polynomial:

$$g(x_1,...,x_n) \approx g(x_1^0,...,x_n^0) + \sum_{r=1}^{n} \frac{\partial g}{\partial x_r}(x_1^0,...,x_n^0)(x_r - x_r^0)$$

If all nonlinear constraints and, if necessary, the objective function are linearized around the point $(x_1^0, \ldots, x_n^0)$ and the optimal solution of the resulting linear programming problem is $(x_1^1, \ldots, x_n^1)$, then the original problem is linearized again around this solution and the new linear programming problem is solved. In this way, an iteration procedure is obtained, and it is known from the theory of nonlinear programming that the solutions of the successive linear programming problems converge to the solution of the original nonlinear optimization problem

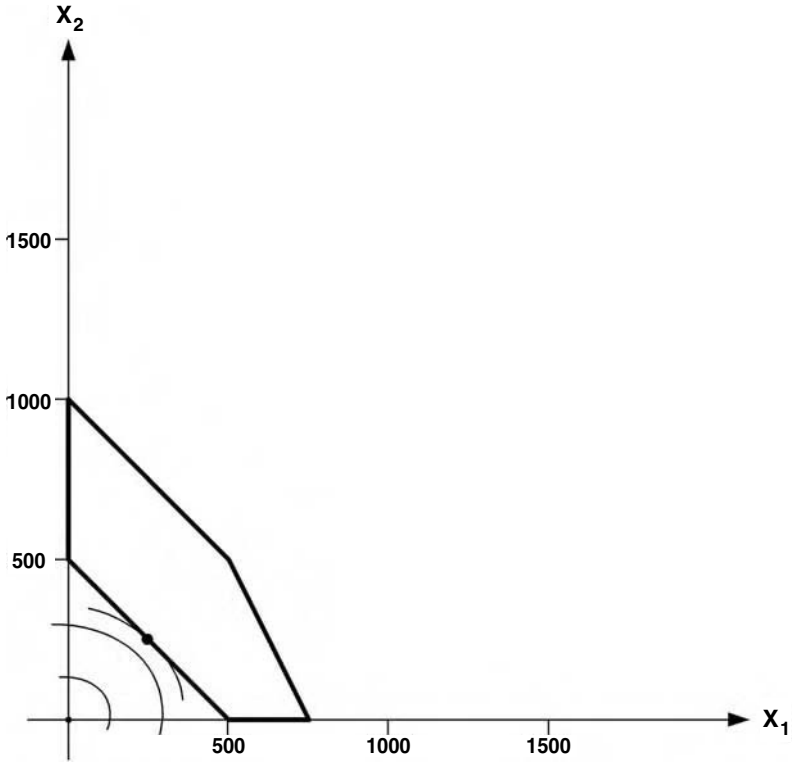


FIGURE 2.2 Illustration of nonlinear optimization.

under reasonable conditions. For more details of nonlinear programming, see, for example, Hadley (1964) or Mangasarian (1962).

2.3 MULTIPLE-CRITERIA OPTIMIZATION

As in the previous session, let X denote the feasible decision space, and let $(f_1, \ldots, f_I)$ be the criteria. Set X reflects the decisions that can be made; however, in the presence of multiple criteria we need to represent the set of all possible outcomes. The set:

$$H = \left\{(f_1(x), \cdots, f_I(x)) | x \in X\right\} \tag{2.15}$$

is called the *criteria space* and shows what we can get by selecting different alternatives.

Example 2.3

Assume that selection has to be made from four available technologies to perform a certain task. These technologies are evaluated by their costs and convenience of

TABLE 2.1
Data for Discrete Problem

Technology	Cost	Convenience
1	–3.0	70
2	–3.5	90
3	–3.8	95
4	–3.2	85

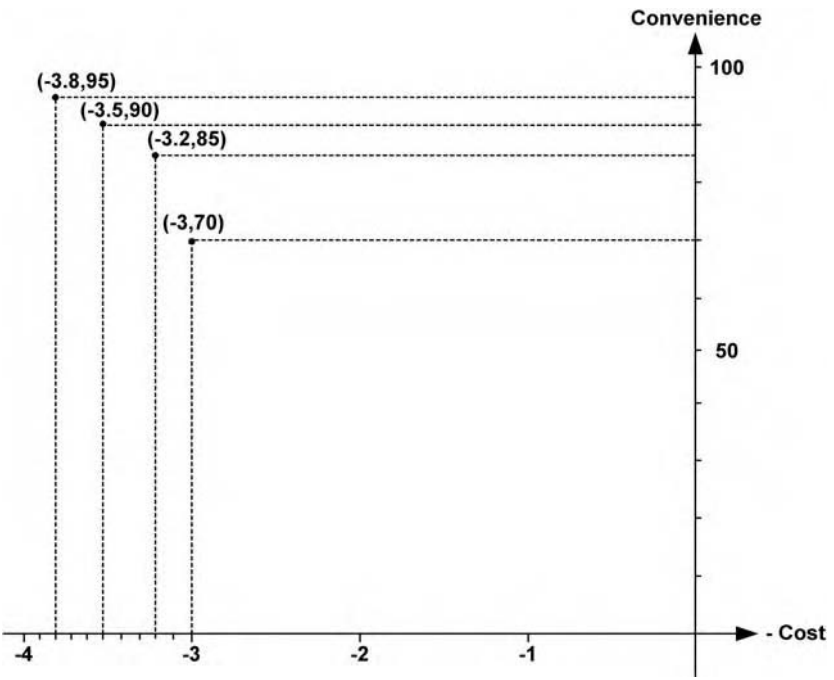


FIGURE 2.3 Criteria space in the discrete case.

usage. The cost data are given in $1000 units, and the convenience of usage is measured in a subjective scale between 0 and 100, where 100 is the best possible measure. The data are given in Table 2.1. We multiplied the cost data by –1 in order to maximize both criteria.

Similar to the single-criterion case, the criteria are sometimes called the *objective functions*. The criteria space has four isolated points, in this case as shown in Figure 2.3. It is clear that none of the points dominates any other point, as lower cost is always accompanied by lower convenience; that is, increasing any objective should result in decreasing the other one. Therefore, any one of the four technologies could be accepted as the solution.

Example 2.4

Consider now a modified version of our earlier first example. Assume that two pollutants are removed by a combination of three alternative technologies. These technologies remove 3, 2, and 1 g/m^3 of the first pollutant and 2, 1, and 3 g/m^3 of the second pollutant. The total amount of the two pollutants removed is:

$$3x_1 + 2x_2 + 1(1000 - x_1 - x_2) = 2x_1 + x_2 + 1000$$

and

$$2x_1 + x_2 + 3(1000 - x_1 - x_2) = -x_1 - 2x_2 + 3000$$

If no condition is given on the usage of the third technology, no constraint on the minimum amount to be removed exists, and no cost is considered, then this problem can be mathematically formulated as:

$$\begin{aligned} &\text{Maximize} && f_1(\underline{x}) = 2x_1 + x_2,\ f_2(\underline{x}) = -x_1 - 2x_2 \\ &\text{Subject to} && x_1, x_2 \geq 0 \\ & && x_1 + x_2 \leq 1000. \end{aligned} \tag{2.16}$$

The feasible decision space is shown in Figure 2.4.

The criteria space can be determined as follows. Because:

$$f_1 = 2x_1 + x_2 \text{ and } f_2 = -x_1 - 2x_2.$$

we have

$$x_1 = \frac{2f_1 + f_2}{3} \text{ and } x_2 = \frac{-f_1 - 2f_2}{3}.$$

The constraints $x_1, x_2 \geq 0$ can be rewritten as:

$$\frac{2f_1 + f_2}{3} \geq 0 \text{ and } \frac{-f_1 - 2f_2}{3} \geq 0$$

that is,

$$2f_1 + f_2 \geq 0 \text{ and } f_1 + 2f_2 \leq 0.$$

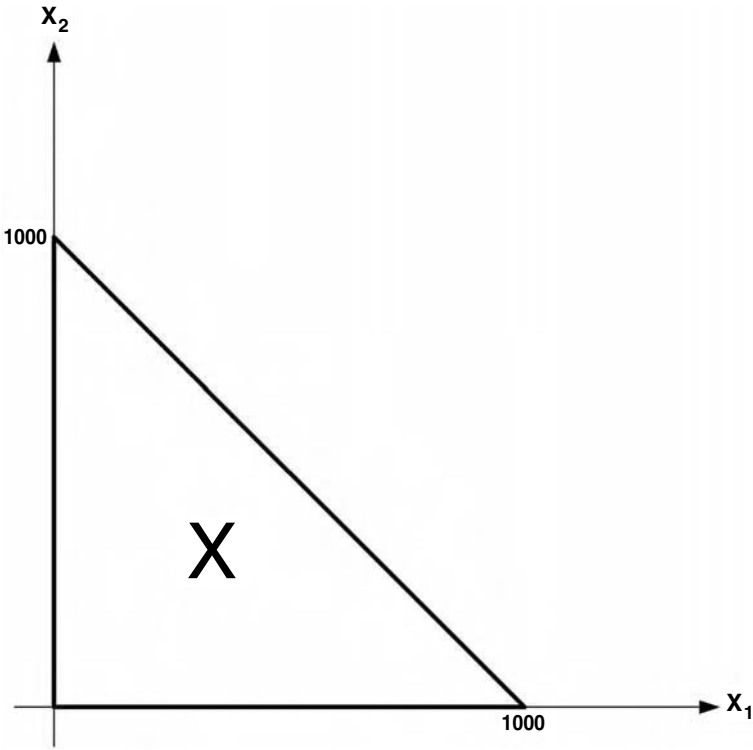


FIGURE 2.4 Feasible decision space for Example 2.4.

The third constraint, $x_1 + x_2 \leq 1000$ now has the form:

$$\frac{2f_1 + f_2}{3} + \frac{-f_1 - 2f_2}{3} \leq 1000$$

that is, $f_1 - f_2 \leq 3000$. Hence, set H is characterized by the following inequalities:

$$\begin{aligned} 2f_1 + f_2 &\geq 0 \\ f_1 + 2f_2 &\leq 0 \\ f_1 - f_2 &\leq 3000 \end{aligned} \tag{2.17}$$

and is illustrated in Figure 2.5. Notice that H is the triangle with vertices (0,0), (1000, –2000), and (2000,–1000), and the points on the linear segment connecting the vertices (0,0) and (2000,–1000) cannot be dominated by any point of H because none of the objectives can be improved without worsening the other. This, however, does not hold for the other points of H when both objectives can be improved simultaneously. Therefore, all points between (0,0) and (2000,–1000) can be accepted as the best solutions.

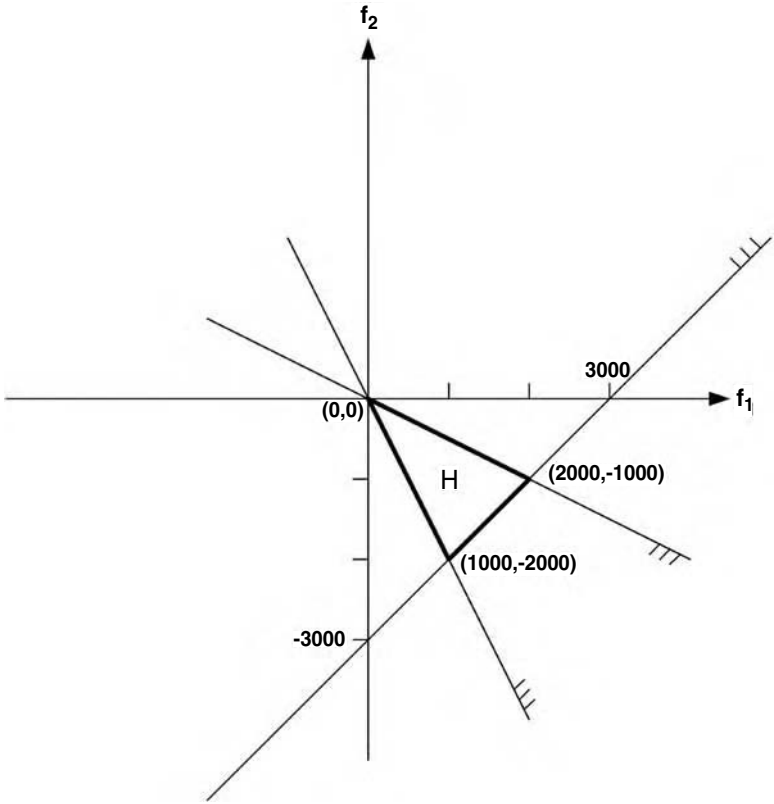


FIGURE 2.5 Criteria space for Example 2.4.

Any point of $\underline{f} \in H$ is called *dominated* if there is another point $\underline{f}^1 \in H$ such that $\underline{f}^1 \geq \underline{f}$ and there is strict inequality in at least one component. Similarly, a point $\underline{f} \in H$ is called *nondominated* if there is not another $f_1 \in H$. In multiple-criteria decision making, nondominated points serve as the solutions of the problem. In the case of single-objective optimization problems, optimal solutions are selected. Optimal solutions are always better than any nonoptimal solution, and in the case of several optimal solutions they are equivalent to each other in the sense that they give the same objective function value. In the case of nondominated solutions, these properties no longer are valid. Consider Figure 2.5, where the point (0,0) is nondominated but is no better than the point (1000, –2000), which is dominated, because $1000 > 0$, thus providing higher value in the first objective function. Also, the points (0,0) and (2000,–1000) are both nondominated and differ in both objective values.

If a single-objective optimization problem has multiple optimal solutions, it does not matter which one is selected, as they give the same objective function value. In multiple-criteria problems, the different nondominated solutions give different objective function values, so it is very important to decide which one of them is chosen, as they lead to different outcomes. To single out a special point from the set of

nondominated solutions, additional preference information is needed from the decision makers to decide on trade-offs between the objectives. For example, in Figure 2.5 both (0,0) and (2000,–1000) are nondominated. We can say that (0,0) is better than (2000,–1000) if a 1000-unit loss in the second objective is not compensated by a 2000-unit gain in the first objective. This kind of decision cannot be made based solely on the concepts discussed before. Depending on the additional preference and/or trade-off information obtained from the decision-makers, different solution concepts and methods are available. These topics are the subjects of the following discussions.

2.3.1 Sequential Optimization

This method is based on the ordinal preference order of the objectives. Assume that the objectives are numbered so that f_I is the most important, f_2 is the second most important, and f_1 is the least important. In applying the method, first f_1 is optimized. If the optimal solution is unique, then the procedure terminates regardless of the values of the less important other objectives. Otherwise, f_1 is kept on its optimal level and f_2 is optimized. If there is a unique optimal solution, then it is selected as the solution of the problem; otherwise, both f_1 and f_2 are kept optimal, and f_3 is optimized, and so on, until there is a unique optimal solution or f_1 is already optimized.

Example 2.5

Consider again the problem given in Example 2.3 and represented by Table 2.1. If cost is more important than convenience, then the least expensive alternative, which is technology 1, is selected. If convenience is more important, then the third technology alternative is chosen. Note that the other two nondominated technologies, alternatives 2 and 4, cannot be chosen by this method.

Example 2.6

In the case of Example 2.4 and Figure 2.5 we have a similar situation. If f_1 is more important than f_2, then point (2000,–1000) is the choice with decision variables:

$$x_1 = \frac{2f_1 + f_2}{3} = 1000 \text{ and } x_2 = \frac{-f_1 - 2f_2}{3} = 0$$

If f_2 is the more important objective, then the point (0,0) is the choice, and the decision:

$$x_1 = \frac{2f_1 + f_2}{3} = 0 \text{ and } x_2 = \frac{-f_1 - 2f_2}{3} = 0$$

is made, in which case only the third alternative pollutant removal technology is used.

2.3.2 The ε-Constraint Method

In the case of sequential optimization, the procedure often terminates before all objectives are optimized. Therefore, less important objectives may have very unfavorable values. In order to avoid such possibilities, the following method can be utilized. Assume that the most important objective is specified, and minimal acceptable levels are given for all other objectives. If f_1 is the most important objective, and (ε_2, ε_3, …, ε_I) are the minimum acceptable levels, then the solution is obtained by solving the following single-objective optimization problem:

$$\begin{aligned} &\text{Maximize} && f_1(x) \\ &\text{Subject to} && x \in X \\ & && f_2(x) \geq \varepsilon_2 \\ & && \vdots \\ & && f_I(x) \geq \varepsilon_I \end{aligned} \tag{2.18}$$

Example 2.7

In the case of Example 2.3, assume that cost is more important than convenience, but we want to have at least 90 in convenience. This means that technologies 1 and 4 are not feasible, but technologies 2 and 3 both satisfy these minimal requirements. The less expensive among these two remaining alternatives is technology 2, so that is our choice.

Example 2.8

Assume that in Example 2.4, f_1 is the more important objective, but we require that $f_2 \geq -500$. With this additional constraint, the criteria space is further reduced to its subset, which has only such points that satisfy this additional condition. The reduced set is shown in Figure 2.6. If f_1 is maximized on the reduced criteria space, then the point (1000,–500) is obtained with decision variables:

$$x_1 = \frac{2f_1 + f_2}{3} = 500 \quad \text{and} \quad x_2 = \frac{-f_1 - 2f_2}{3} = 0$$

That is, pollutant removal technologies 1 and 3 are used at an equal rate.

2.3.3 The Weighting Method

When applying this method, all objectives are taken into account by importance weights supplied by the decision makers. Assume that (c_1, c_2, …, c_I) are the relative importance factors of the objectives which are specified by the decision makers on

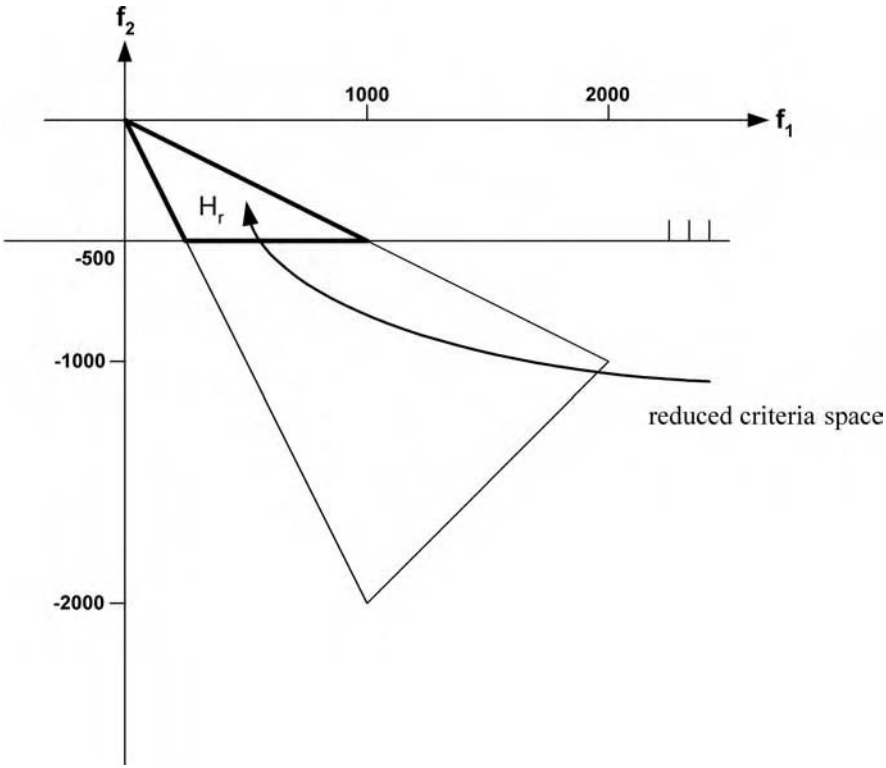


FIGURE 2.6 Reduced criteria space for Example 2.8.

some subjective basis or are obtained by some rigorous method such as pair-wise comparisons. It is assumed that $c_i > 0$ for all i, and

$$\sum_{i=1}^{I} c_i = 1 .$$

When applying the weighting method, the composite objective is a weighted average of all objectives, where the weights are the relative importance factors (c_1, …, c_I). Hence, the single objective optimization problem:

$$\begin{aligned} &\text{Maximize} \quad c_1 f_1(x) + c_2 f_2(x) + \ldots + c_I f_I(x) \\ &\text{Subject to} \quad x \in X \end{aligned} \tag{2.19}$$

is solved. The method in this form presents a major difficulty. First, the different objectives have different units. For example, cost is measured in dollars, convenience

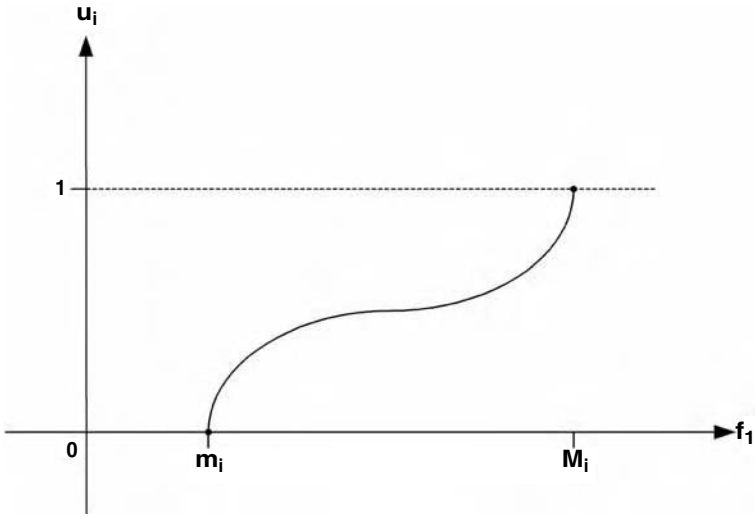


FIGURE 2.7 Utility function of objective f_i.

is given as a subjective measure, pollutant removal is given in g/m^3, and so on. Therefore, the composite objective has no practical meaning, as we are adding apples and oranges. The second difficulty is based on the observation that, by changing the unit of any of the objectives, we multiply its weight by the conversion constant so the optimal solution usually changes. Both difficulties can be overcome by transforming all objectives into a common dimensionless satisfaction scale. This transformation is called the *normalizing* procedure, and it can be done by a simple linear transformation or by using special utility functions supplied by the decision makers.

The simple linear normalizing procedure is outlined first. Let f_i denote an objective function. Let m_i and M_i denote its smallest and largest possible values, respectively. These values may be computed by simply minimizing and maximizing f_i over the feasible decision space X, or their values can be supplied by the decision makers based on subjective judgments. The normalized objective function can then be obtained as:

$$\overline{f}_i(x) = \frac{f_i(x) - m_i}{M_i - m_i} \tag{2.20}$$

In many practical cases, the decision makers are able to provide a utility function to each objective that shows their satisfaction levels for the different values of that objective. Such a function is illustrated in Figure 2.7, where m_i gives the least satisfaction level (0) and M_i gives the highest satisfaction level (1, or 100%). Then, objective f_i is replaced by the corresponding utility value:

$$\overline{f}_i(x) = u_i\left(f_i(x)\right) \tag{2.21}$$

Notice that Eq. (2.20) corresponds to the linear utility function:

$$u_i(f_i) = (f_i - m_i) / (M_i - m_i)$$

We always require that the lowest possible utility value be 0 and the highest 1, so $\bar{f}_i(x) \in [0,1]$ for all $x \in X$. Then, Eq. (2.19) can be modified:

$$\begin{aligned} &\text{Maximize} \quad c_1\bar{f}_1(x) + c_2\bar{f}_2(x) + \ldots + c_I\bar{f}_I(x) \\ &\text{Subject to} \quad x \in X \end{aligned} \tag{2.22}$$

In this new composite objective, only dimensionless values are added, and the linear combination of the normalized objectives can be interpreted as an average satisfaction level based on all objectives.

Example 2.9

Assume that in Example 2.3 the decision makers consider \$4000 as the worst possible cost, and they believe that a cost of \$3000 would give them 100% satisfaction. Then $m_1 = -4$, and $M_1 = -3$, and, if linear utility function is assumed between these values, then the linearization is based on the linear transformation:

$$\bar{f}_1 = \frac{f_1 + 4}{-3 + 4}$$

so the cost values in Table 2.1 must be modified accordingly. The convenience data are already given in satisfaction levels (in %), so these numbers are divided by 100 to transform them into the unit interval [0,1]. The normalized table is provided in Table 2.2.

TABLE 2.2
Normalized Objectives for Discrete Problem

Technology	Cost	Convenience
1	1.0	0.70
2	0.5	0.90
3	0.2	0.95
4	0.8	0.85

Assume that equal importance is given to the objectives; that is $c_1 = c_2 = 0.5$. Then, the composite objective values for the four technology variants are as follows:

$$0.5(1) + 0.5(0.7) = 0.85$$

$$0.5(0.5) + 0.5(0.9) = 0.70$$

$$0.5(0.2) + 0.5(0.95) = 0.575$$

$$0.5(0.8) + 0.5(0.85) = 0.825.$$

Thus, the first alternative is the choice.

Example 2.10

Consider again the problem outlined in Example 2.4. From Figure 2.5, we see that the smallest values of the objectives are $m_1 = 0$ and $m_2 = -2000$; the largest values are $M_1 = 2000$ and $M_2 = 0$. If the simple normalization is performed, then the new objectives are:

$$\bar{f}_1(x) = \frac{f_1(x)}{2000} \quad \text{and} \quad \bar{f}_2(x) = \frac{f_2(x) + 2000}{2000}$$

The modified criteria space can be obtained by transforming the vertices and constructing the convex hull generated by the new vertices. Notice that by using the above transformation we have:

$$(0,0) \mapsto (0,1)$$
$$(1000,-2000) \mapsto (0.5, 0)$$
$$(2000,-1000) \mapsto (1, 0.5)$$

These new vertices and the transformed criteria space $\overline{H}$ are illustrated in Figure 2.8. Assume next that equal importance is assigned to the objectives $c_1 = c_2 = 0.5$. Then, we minimize $0.5\bar{f}_1 + 0.5\bar{f}_2$ over the set $\overline{H}$. It is easy to see that the vertex (1,0.5) gives the highest composite objective value. The values of the original objectives can be easily obtained by solving the following equations:

$$\bar{f}_1 = \frac{f_1}{2000} = 1 \quad \text{and} \quad \bar{f}_2 = \frac{f_2 + 2000}{2000} = 0.5$$

which yields $f_1 = 2000$ and $f_2 = -1000$. The decision variables are then obtained:

$$x_1 = \frac{2f_1 + f_2}{3} = 1000 \quad \text{and} \quad x_2 = \frac{-f_1 - 2f_2}{3} = 0$$

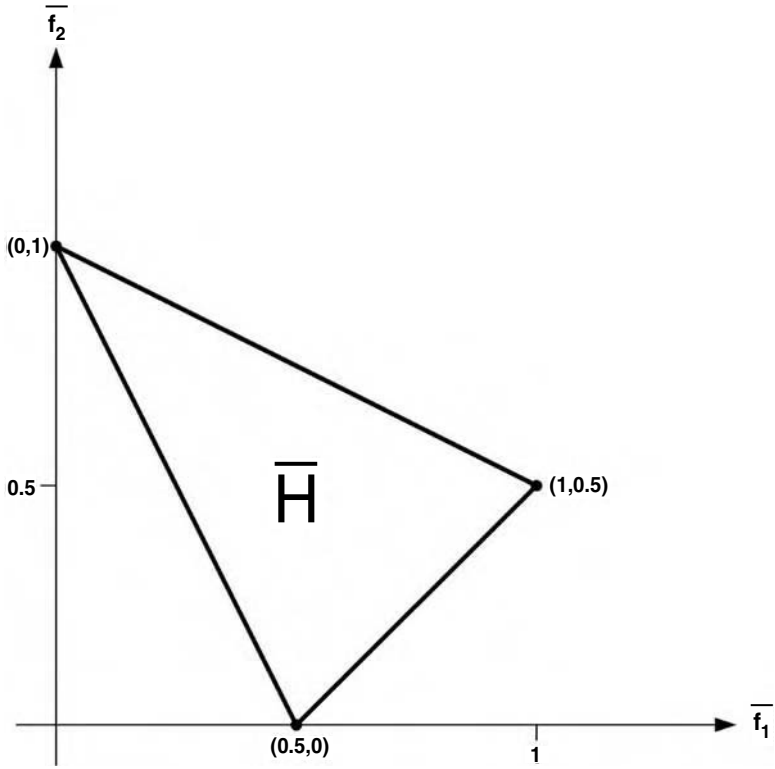


FIGURE 2.8 Transformed criteria space for Example 2.4.

These equations for x_1 and x_2 were defined earlier in Example 2.4.

2.3.4 Direction-Based Method

In many practical cases, the decision makers are able to specify a point in the criteria space (which might not be feasible in some cases) and a direction in which the objectives have to be improved. The given point might be the current state of the system that has to be improved, or it can be some subjective worst possibility. If $\underline{f}_*$ is the given point and $\underline{\vee} > \underline{0}$ is the direction of improvement, then this concept can be mathematically modeled as the following single-objective optimization problem:

$$\begin{aligned} &\text{Maximize} && t \\ &\text{Subject to} && \underline{f}_* + t\underline{\vee} \in H \end{aligned} \tag{2.23}$$

The graphical representation of this method is shown in Figure 2.9. Starting from point $\underline{f}_*$, we move in direction of $\underline{\vee}$ as far as possible. This method has two major difficulties. First, the obtained solution might be dominated, so the result cannot be accepted in such cases. Figure 2.10 shows this possibility. Second, this method

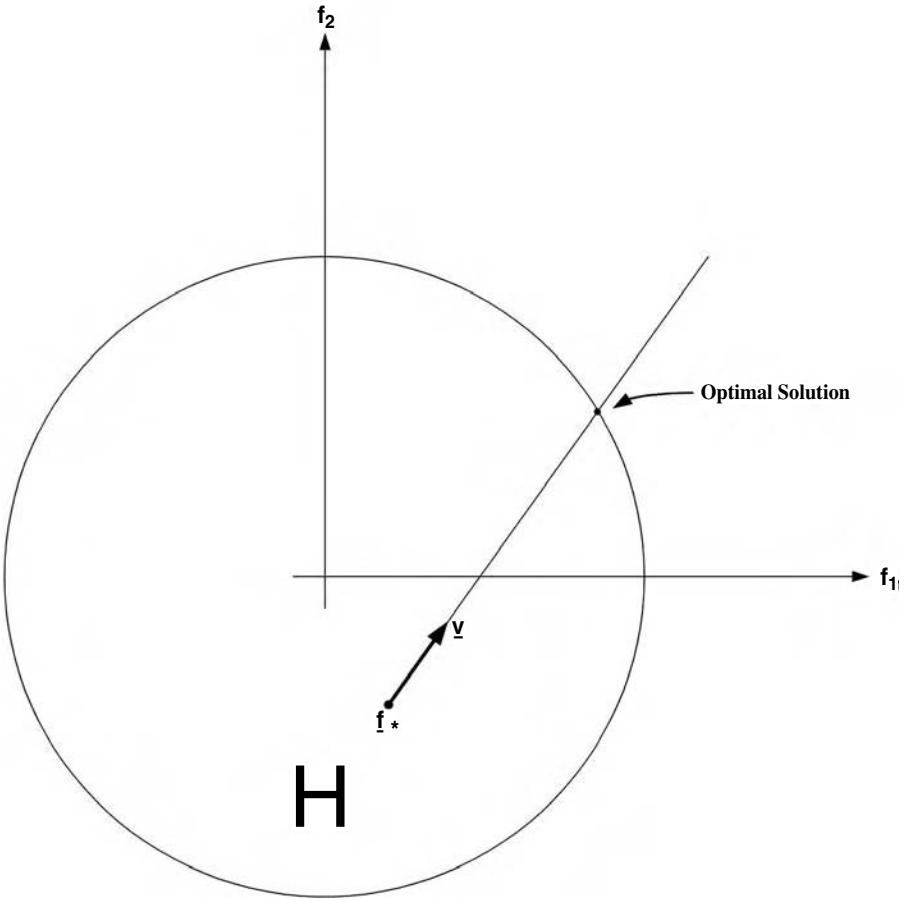


FIGURE 2.9 Illustration of the direction-based method.

cannot be applied for discrete problems, as there is usually no other feasible point along the path we move in the improvement.

Example 2.11

Consider again the criteria space shown in Figure 2.5. Assume $\underline{f}_*$ is selected as the vector having the smallest values of the objectives in the components; $\underline{f}_*$ = (0,–2000). If the two objectives have to be improved equally, then $\underline{v}$ = (1,1). The graphical solution of problem (2.23) is shown in Figure 2.11. The solution is the intercept of the two lines with equations:

$$f_2 = f_1 - 2000 \text{ and } f_1 + 2f_2 = 0$$

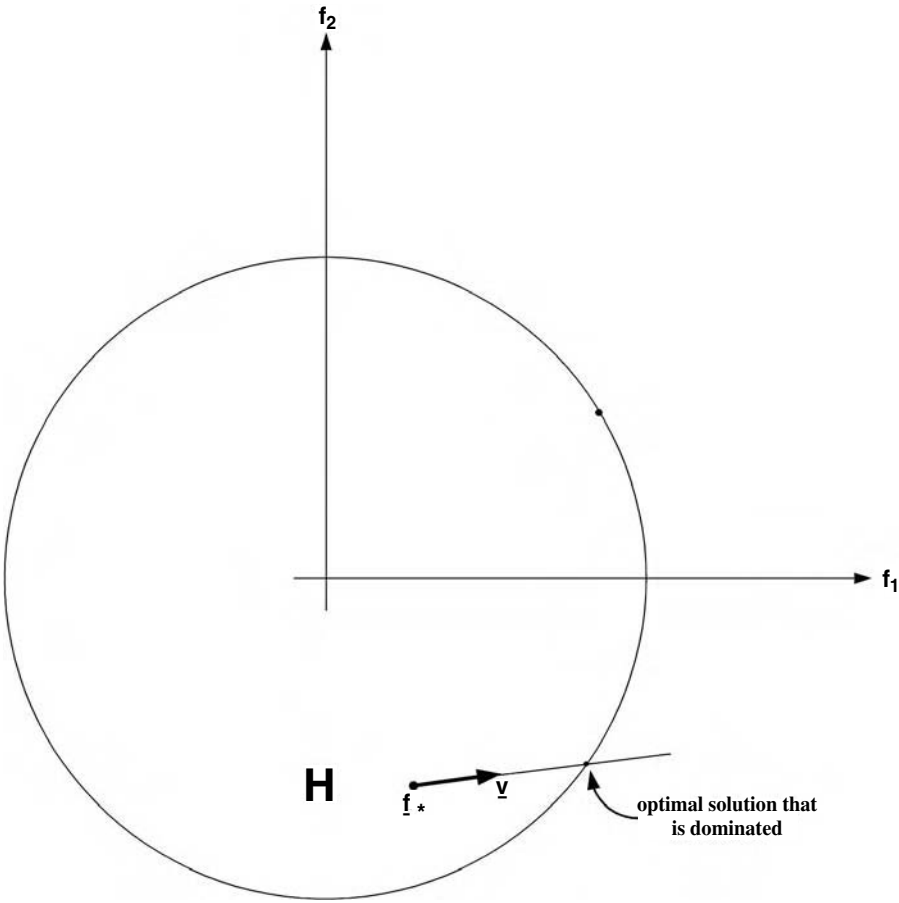


FIGURE 2.10 Illustration of the direction-based method.

It is easy to see that:

$$f_1 = \frac{4000}{3} \quad \text{and} \quad f_2 = -\frac{2000}{3}$$

and the decision variables are:

$$x_1 = \frac{2f_1 + f_2}{3} \approx 666.67 \quad \text{and} \quad x_2 = \frac{-f_1 - 2f_2}{3} = 0$$

Direction-based methods are sometimes applied when an ideally best point is known, which is not feasible. In that case, we start relaxing the values of the objective functions in a given direction until a feasible point is reached. Let $\underline{f}^*$ denote the

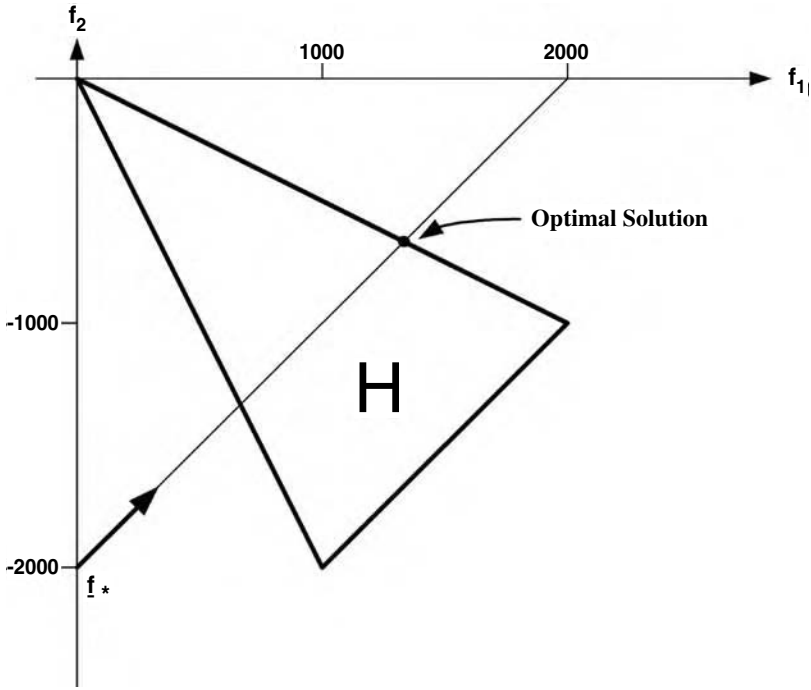


FIGURE 2.11 Solution by a direction-based method.

given point and $\underline{\vee} > \underline{0}$ be the given direction of relaxation. Equation (2.23) is now modified as:

$$\begin{aligned} &\text{Minimize} \quad t \\ &\text{Subject to} \quad \underline{f}^* - t\underline{\vee} \in H \end{aligned} \tag{2.24}$$

The graphical representation of this method is shown in Figure 2.12.

Example 2.12

In the case of the previous example select $\underline{f}^*$ = (2000,0) and $\underline{\vee}$ = (1,1). Then, the line of relaxation coincides with the line of improvement, so the optimal solution is the same as in the previous example.

2.3.5 Distance-Based Methods

Assume now that an ideally best point $\underline{f}^*$ is computed or obtained from the decision makers. If $\underline{f}^*$ is feasible, then the corresponding decision variables are obtained by solving the equations:

$$\underline{f}(x) = \underline{f}^*$$

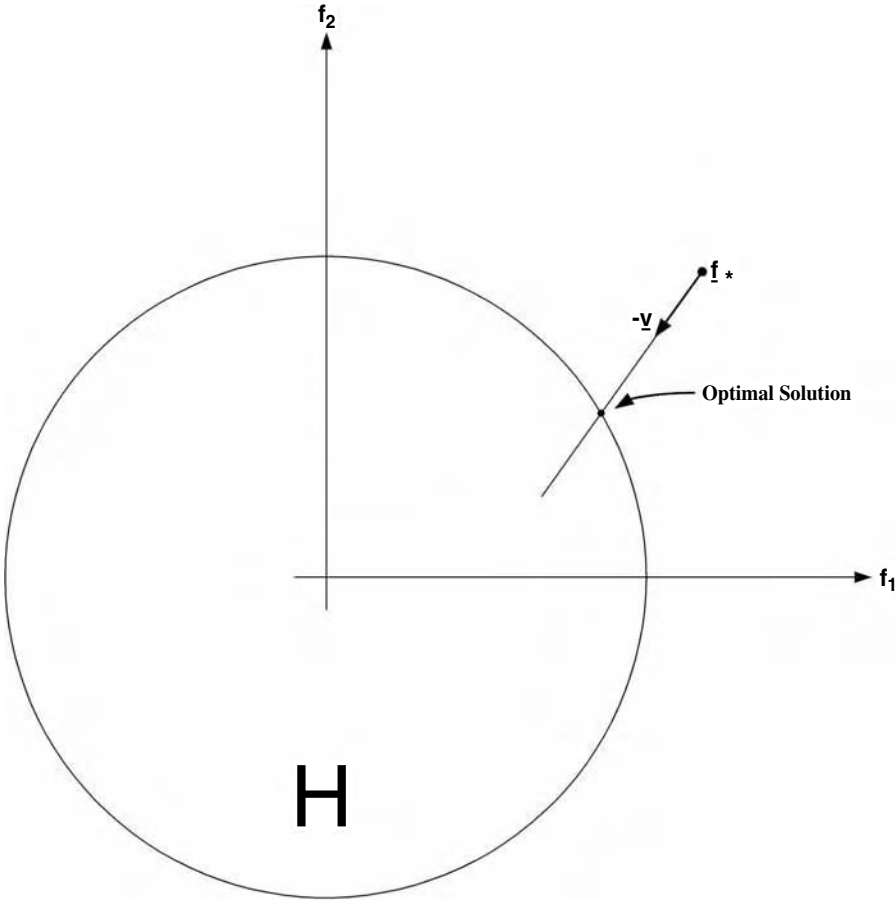


FIGURE 2.12 Illustration of problem (2.24).

Otherwise, we are looking for the feasible point that has minimal distance from the ideal point. If $\rho(\underline{f}(x), \underline{f}^*)$ denotes the distance between points $\underline{f}(x)$ and $\underline{f}^*$, then this concept can be mathematically modeled as the single-objective optimization problem:

$$\begin{aligned} &\text{Minimize} && \rho(\underline{f}(x), \underline{f}^*) \\ &\text{Subject to} && x \in X \end{aligned} \tag{2.25}$$

The graphical illustration of this method is very similar to Figure 2.12, where the optimal solution has the smallest distance from the ideal point. In practical applications, one of the following distances is usually used:

$$\rho_1(\underline{f}(x), \underline{f}^*) = \sum_{i=1}^{I} c_i \left| f_i(x) - f_i^* \right| \tag{2.26}$$

$$\rho_2(\underline{f}(x), \underline{f}^*) = \sqrt{\sum_{i=1}^{I} c_i (f_i(x) - f_i^*)^2} \tag{2.27}$$

and

$$\rho_\infty(\underline{f}(x), \underline{f}^*) = \max_i \left\{ c_i \left| f_i(x) - f_i^* \right| \right\} \tag{2.28}$$

where $c_1, c_2, \ldots, c_I$ are given positive weights and $f_1^*, f_2^*, \ldots, f_I^*$ are the components of $\underline{f}^*$. The ρ_1 distance represents full compensation among the objectives, as a unit increase in $| f_i(x) - f_i^* |$ can be compensated by c_i/c_j units of $| f_j(x) - f_j^* |$ in order to have the same distance. In the case of distance, ρ_2 partial compensation is assumed. The compensating amounts are proportional to the quantities, and in the case of ρ_∞ no compensation is assumed, as only the largest value of $c_i | f_i(x) - f_i^* |$ affects the distance. Smaller values do not count, so they can have any smaller values without changing the distance. Because weights are included in the distance functions, the objective functions are usually normalized in a way similar to the weighting method.

Example 2.13

Consider again the normalized discrete problem of Example 2.9. Assume $\underline{f}^* = (1,1)$ and equal weights are selected. By selecting the ρ_1 distance we have:

$$\rho_1(\underline{f}(1), \underline{f}^*) = 0.5|1-1| + 0.5|1-0.7| = 0.15$$

$$\rho_1(\underline{f}(2), \underline{f}^*) = 0.5|1-0.5| + 0.5|1-0.9| = 0.3$$

$$\rho_1(\underline{f}(3), \underline{f}^*) = 0.5|1-0.2| + 0.5|1-0.95| = 0.425$$

$$\rho_1(\underline{f}(4), \underline{f}^*) = 0.5|1-0.8| + 0.5|1-0.85| = 0.175$$

The smallest distance occurs in the first case, so the first alternative is selected.

Example 2.14

We will consider now the normalized criteria space shown in Figure 2.8. Let the ideal point be selected again as $\underline{f}^* = (1,1)$, and let us choose the ρ_∞ distance with equal weights. In this case, we have the following nonlinear programming problem:

$$\text{Minimize} \quad \max\left\{\left|1-\frac{2x_1+x_2}{2000}\right|, \left|1-\frac{2000-x_1-2x_2}{2000}\right|\right\}$$

$$\text{Subject to} \quad x_1, x_2 \geq 0$$
$$x_1 + x_2 \leq 1000$$

where we used the normalizing equations derived in Example 2.10. This is a nonlinear optimization problem; however, we are able to rewrite it as a higher dimensional linear programming problem. If we introduce the variable M for the objective function, then:

$$\left|1-\frac{2x_1+x_2}{2000}\right| \leq M \text{ and } \left|1-\frac{2000-x_1-2x_2}{2000}\right| \leq M$$

We can eliminate the absolute values by having:

$$-M \leq 1-\frac{2x_1+x_2}{2000} \leq M \text{ and } -M \leq 1-\frac{2000-x_1-2x_2}{2000} \leq M$$

These inequalities can be further simplified as:

$$2x_1 + x_2 - 2000M \leq 1$$
$$2x_1 + x_2 + 2000M \geq 1$$
$$x_1 + 2x_2 + 2000M \geq -1$$
$$x_1 + 2x_2 - 2000M \leq -1$$

Therefore, the above nonlinear optimization problem is equivalent to the three-dimensional linear programming problem:

$$\text{Minimize} \quad M$$

$$\text{Subject to} \quad x_1, x_2 \geq 0$$
$$x_1 + x_2 \leq 1000$$
$$2x_1 + x_2 - 2000M \leq 1$$
$$2x_1 + x_2 + 2000M \geq 1$$
$$x_1 + 2x_2 + 2000M \geq -1$$
$$x_1 + 2x_2 - 2000M \leq -1$$

The application of the simplex method gives the optimal solution:

$$x_1 = \frac{6000}{9} \approx 666.67 \text{ and } x_2 = 0$$

In some applications, an ideally worst point is given and the feasible point that is as far as possible from this point in the criteria space is selected. If $\underline{f}_*$ denotes this point, then this concept can be mathematically modeled as follows:

$$\text{Maximize} \quad \rho(\underline{f}(x), \underline{f}_*)$$

$$\text{Subject to} \quad x \in X \tag{2.29}$$

where ρ is any distance of I-dimensional vectors.

Finally, we mention that a comprehensive summary of multiple-criteria decision-making methods can be found, for example, in Szidarovszky et al. (1986).

2.4 GROUP DECISION MAKING

In this subsection, we assume multiple decision makers, with each decision maker having a single objective function to maximize. One way of solving such problems is to consider the problem as a multiple-criteria decision problem, where the objectives of the different decision makers form the multiple objective of the problem. If the decision makers have equal importance, then equal weights are selected; otherwise, the importance, or powers, of the decision makers are the weights.

Another approach is based on social choice procedures. Assume that I decision makers have to choose a commonly acceptable alternative from a finite set. Assume also that each decision maker has his own ranking of the alternatives. The data then can be presented in a table, where the rows correspond to the decision makers and the columns show the decision alternatives. Each row of the table represents the rankings of the decision maker, so each row has a permutation of the integers 1, 2, …, n, where n is the number of alternatives. Such a table is shown in Table 2.3. The last column gives the weights of the decision makers. We assume again that:

$$c_i \geq 0 \text{ and } \sum_{i=1}^{I} c_i = 1$$

In the case of *plurality voting*, we compute for each alternative the total weight for those decision makers who consider the alternative to be the best, and the alternative with the largest such number is selected as the common choice. Mathematically, this method can be described as follows. For each alternative, we first compute:

$$A_k = \sum_{i=1}^{I} c_i f(a_{ik}) \quad (k = 1, \ldots, n) \tag{2.30}$$

TABLE 2.3
Input for Social Choice

	Alternatives				
Decision Makers	**1**	**2**	**…**	***n***	**Weights**
1	a_{11}	a_{12}	…	a_{1n}	c_1
2	a_{21}	a_{22}	…	a_{2n}	c_2
⋮	⋮	⋮	…	⋮	⋮
I	a_{I1}	a_{I2}	…	a_{In}	c_I

where

$$f(x)=\begin{cases}1 & \text{if } a_{ik}=1\\ 0 & \text{otherwise,}\end{cases}$$

and then alternative k_0 is selected if:

$$A_{k_0}=\max_k\{A_k\} \tag{2.31}$$

Notice that this method considers only first rankings.

The *Borda count* takes all rankings into consideration. For each alternative we first compute the average weighted ranking of each alternative, and the one with the smallest average value is selected as the social choice. So, in the first step we compute:

$$B_k=\sum_{i=1}^{I} c_i a_{ik} \tag{2.32}$$

for $k = 1, 2, \ldots, n$. Alternative k_0 is the social choice if:

$$B_{k_0}=\min_k\{B_k\} \tag{2.33}$$

The *Hare system* is based on the successive deletion of less desirable alternatives. First, we check to see if an alternative is selected as being the best with total weights of at least 0.5; this alternative is the social choice, and the procedure terminates. Otherwise, the alternative with the smallest such total weight is deleted, the table is adjusted, and we go back to the first step and check again. If one among the remaining alternatives has a total weight of at least 0.5 for its rankings of being best, then this alternative is the social choice. Otherwise, we delete an alternative again. This procedure continues until in the first step we have a clear choice. First we compute $A_1, \ldots, A_n$ by using Eq. (2.30). If, for a k, $A_k \geq 0.5$, then alternative k becomes the social choice. Otherwise, let:

$$A_{k_0} = \min_k \{A_k\} \tag{2.34}$$

Then, we delete alternative k_0 from the table, so n becomes $n - 1$, and we modify the a_{ik} values as follows:

$$a_{ik} = \begin{cases} a_{ik} & \text{if} \quad a_{ik} < a_{ik_0} \\ a_{ik} - 1 & \text{if} \quad a_{ik} > a_{ik_0} \end{cases} \tag{2.35}$$

Then, we go back to the first step, computing again the new A_k numbers.

When applying *pair-wise comparisons*, the decision makers first have to agree on the order of comparisons. Let $k_1, k_2, \ldots, k_n$ be the order, which is a permutation of integers 1, 2, …, n. Alternatives k_1 and k_2 are compared first by computing the total weight of the decision makers who prefer alternative k_1 to k_2. If this total weight is at least 0.5, then k_1 is considered better than k_2; otherwise, k_2 is considered better than k_1. Then, the winner is compared to k_3. The winner of this comparison is compared to k_4, and so on. The winner of the last comparison is considered to be the social choice. When comparing two alternatives k and l we compute the value $N(k,l)$. Alternative k is considered better than l if $N(k,l) \geq 0.5$:

$$N(k,l) = \sum_i \left\{ c_i \middle| a_{ik} < a_{il} \right\} \tag{2.36}$$

Example 2.15

Assume that five decision makers have to choose the best design plan for a water resources project. The data are given in Table 2.4. The last column shows that decision makers 1 and 2 have higher priorities than the third, who is more important than the fourth and fifth decision makers. In applying plurality voting, we first compute:

$$A_1 = 0.3, A_2 = 0.4, A_3 = 0, A_4 = 0.3$$

Because A_2 is the largest value, the second plan is considered to be the social choice.

For a Borda count we compute the B_k values:

$$B_1 = 3(0.3) + 1(0.3) + 4(0.2) + 2(0.1) + 4(0.1) = 2.6$$

$$B_2 = 1(0.3) + 4(0.3) + 3(0.2) + 3(0.1) + 1(0.1) = 2.5$$

$$B_3 = 2(0.3) + 3(0.3) + 2(0.2) + 4(0.1) + 2(0.1) = 2.5$$

$$B_4 = 4(0.3) + 2(0.3) + 1(0.2) + 1(0.1) + 3(0.1) = 2.4$$

The smallest value occurs at the fourth alternative, so it is the social choice again.

TABLE 2.4
Data for Example 2.15

Decision Makers	Alternatives (1 2 3 4)	Weights
1	3 1 2 4	0.3
2	1 4 3 2	0.3
3	4 3 2 1	0.2
4	2 3 4 1	0.1
5	4 1 2 3	0.1

TABLE 2.5
Modified Data for Example 2.15

Decision Makers	Alternatives (1 2 4)	Weights
1	2 1 3	0.3
2	1 3 2	0.3
3	3 2 1	0.2
4	2 3 1	0.1
5	3 1 2	0.1

When applying the Hare system, we again consider the A_k numbers computed in plurality voting. Because none of these values is at least 0.5, no choice is made. The smallest value occurs at the third alternative, so it is deleted. The modified table is shown in Table 2.5.

We recompute first the A_k values from the modified table:

$$A_1 = 0.3, A_2 = 0.4, A_4 = 0.3$$

None of these values is at least 0.5, so no clear choice is made. A_1 and A_4 are both the smallest, so we delete one of them (say, alternative A_4) from the table. The new table is given in Table 2.6. Because $A_1 = 0.4$ and $A_2 = 0.6$, the second alternative is the social choice.

Assume that in applying pair-wise comparisons the decision makers agree in the order 1, 2, 3, and 4. Alternatives 1 and 2 are compared first:

$$N(1,2) = 0.3 + 0.1 = 0.4 < 0.5$$

so alternative 2 is considered better. Then, alternative 2 is compared to alternative 3:

$$N(2,3) = 0.3 + 0.1 + 0.1 = 0.5$$

TABLE 2.6
Second Modified Table for Example 2.15

Decision Makers	Alternatives (1 2)	Weights
1	2 1	0.3
2	1 2	0.3
3	2 1	0.2
4	1 2	0.1
5	2 1	0.1

which shows that no preference exists between these alternatives. However, in the comparison order the decision makers wanted to bring alternative 3 into the sequence of comparisons later than alternative 2, so they gave higher priority to the third alternative. Therefore, it is compared to the last one:

$$N(3,4) = 0.3 + 0.1 = 0.4 < 0.5$$

Therefore, the fourth alternative is the social choice. The interested reader may find more details of voting and social choice procedures in Taylor (1995).

2.5 CONFLICT RESOLUTION

As we mentioned earlier, the presence of multiple decision makers makes the decision-making process much more difficult, as they usually have different priorities and goals; therefore, reaching a compromise among the conflicting objectives is sometimes very difficult. Several approaches are available to solve conflict situations. One might consider the problem as a multiple-objective optimization problem that addresses the objectives of the different decision makers. Conflict situations can be also modeled as social choice problems in which the rankings of the decision makers are taken into account in the final decision. A third way of resolving conflicts was offered by Nash, who considered a certain set of conditions the solution has to satisfy and proved that exactly one solution satisfies his "fairness" requirements. This section outlines the Nash bargaining solution.

Assume that we have I decision-makers. Let X be the decision space and $f_i: X \mapsto R$ be the objective function of decision maker i. The criteria space is defined as:

$$H = \left\{ \underline{u} \,\middle|\, \underline{u} \in R^I, \underline{u} = (u_i), u_i = f_i(x) \text{ with some } x \in X \right\}$$

It is also assumed that when the decision makers are unable to reach an agreement, all decision makers will get low objective function values. Let d_i denote this value for decision-maker i, and let $\underline{d} = (d_1, d_2, \ldots, d_I)$; therefore, the conflict is completely defined by the pair $(H,\underline{d})$, where H is the set of all possible outcomes

and $\underline{d}$ represents the outcomes if no agreement is reached. Thus, any solution of the conflict depends on both H and $\underline{d}$. Let the solution be denoted as a function of H and $\underline{d}$: $\underline{\varphi}(H,\underline{d})$. It is assumed that the solution function satisfies the following conditions:

1. The solution has to be feasible: $\underline{\varphi}(H,\underline{d}) \in H$.
2. The solution has to provide at least the disagreement outcome to all decision makers: $\underline{\varphi}(H,\underline{d}) \geq \underline{d}$.
3. The solution has to be *nondominated*; that is, there is no $\underline{f} \in H$ such that $\underline{f} \neq \underline{\varphi}(H,\underline{d})$ and $\underline{f} \geq \underline{\varphi}(H,\underline{d})$.
4. The solution must not depend on unfavorable alternatives; that is, if $H_1 \subset H$ is a subset of H such that $\underline{\varphi}(H,\underline{d}) \in H_1$, then $\underline{\varphi}(H,\underline{d}) = \underline{\varphi}(H_1,\underline{d})$.
5. Increasing linear transformation should not alter the solution. Let T be a linear transformation on such that $T(f) = (\alpha_1 f_1 + \beta_1, \ldots, \alpha_I f_I + \beta_I)$, with $\alpha_i > 0$ for all i, then $\underline{\varphi}(T,(H),T(\underline{d})) = T(\underline{\varphi}(H,\underline{d}))$.
6. If two decision makers have equal positions in the definition of the conflict, then they must get equal objective values at the solution. Decision makers i and j have equal position if $d_i = d_j$, and any vector $\underline{f} = (f_1, \ldots, f_I) \in H$ if and only if $(\overline{f}_1, \ldots, \overline{f}_I) \in H$, with $\overline{f}_i = f_j$, $\overline{f}_j = f_i$, $\overline{f}_1 = f_1$, for $l \neq i, j$. Then we require that $\underline{\varphi}_i(H,\underline{d}) = \underline{\varphi}_j(H,\underline{d})$.

Before presenting the method for finding the solution that satisfies these properties, some remarks are in order. The feasibility condition requires that the decision makers cannot get more than the amount available. No decision maker would agree to an outcome that is worse than the amount he would get anyway without the agreement. This property is given in condition 2. The requirement that the solution must be nondominated shows that no better possibility is available for all. The fourth requirement says that if certain possibilities become not feasible but the solution remains feasible, then the solution must not change. If any of the decision makers changes the unit of his objective, then a linear transformation is performed on H. The fifth property requires that the solution must remain the same. The last requirement shows a certain kind of fairness by stating that if two decision makers have the same outcome possibilities and the same disagreement outcome, then there is no reason to distinguish among them in the final solution.

If H is convex, closed, and bounded, and there is at least one $\underline{f} \in H$ such that $\underline{f} > \underline{d}$, then there is a unique solution $\underline{f}^* = \underline{\varphi}(H,\underline{d})$, which can be obtained as the unique solution of the following optimization problem:

$$\begin{aligned} &\text{Maximize} && (f_1 - d_1)(f_2 - d_2)\ldots(f_I - d_I) \\ &\text{Subject to} && f_i \geq d_i \qquad (i = 1,2,\ldots,I) \\ & && \underline{f} = (f_1,\ldots,f_I) \in H \end{aligned} \tag{2.37}$$

The objective function is called the *Nash product*. Notice that this method can be considered as a special distance-based method when the geometric distance is maximized from the disagreement point $\underline{d}$.

Example 2.16

Consider again the discrete problem given earlier in Table 2.1. Assume that cost and convenience are the objectives of two different decision makers. Assume that $\underline{d}$ = (–4,50), where –4 is considered as the worst possibility of cost and 50 is the worst possibility for convenience. The values of the Nash products for the four alternatives are as follows:

$$(-3 + 4)\ (70 - 50) = 20$$

$$(-3.5 + 4)\ (90 - 50) = 20$$

$$(-3.8 + 4)\ (95 - 50) = 9$$

$$(-3.2 + 4)\ (85 - 50) = 28$$

Because the last alternative gives the largest value, the fourth technology variant is considered to be the solution.

Example 2.17

Consider next the continuous example shown in Figure 2.5, and assume that f_1 and f_2 are the objectives of two decision makers, such as might occur when the removal of two kinds of pollutants is controlled by two different officials. Assume that $\underline{d}$ = (0,–2000) is the worst value of the two objectives. In this case, Eq. (2.37) can be rewritten as:

$$\text{Maximize} \quad f_1(f_2 + 2000)$$

$$\text{Subject to} \quad \left(f_1, f_2\right) \in H$$

A simple computer study shows that the optimal solution is:

$$f_1 = 2000 \ \text{ and } \ f_2 = -1000,$$

and the corresponding decision variables are given as:

$$x_1 = \frac{2f_1 + f_2}{3} = 1000, \qquad x_2 = \frac{-f_1 - 2f_2}{3} = 0$$

In many practical cases, conflict resolution is needed when the decision makers have different powers. In such cases, problem (2.37) becomes:

$$\text{Maximize} \quad \left(f_1 - d_1\right)^{c_1}\left(f_2 - d_2\right)^{c_2} \ldots \left(f_1 - d_I\right)^{c_I} \tag{2.38}$$

$$\text{Subject to} \quad f_i \geq d_i \qquad \left(i = 1, 2, ..., I\right)$$
$$\left(f_1, \ldots, f_I\right) \in H$$

Here, $(c_1, c_2, \ldots, c_I)$ show the relative powers of the decision makers. The solution of this problem is usually called the nonsymmetric Nash bargaining solution.

Conflict resolution is a special chapter in game theory, so any text or monograph on game theory is a potential source for further details. See, for example, Forgo et al. (1999).

2.6 SPECIAL METHODOLOGY

2.6.1 Dynamic Programming

In water resources planning, dynamism of the system is a major concern that must be taken into account in modeling and optimization. Dynamic programming is the most popular tool for optimizing dynamic systems.

Let t denote time, where $t = 0, 1, 2, \ldots, T$. Let x_t be the state of the system at time period t, and let u_t denote the decision (or control) at the same time period. It is assumed that the state changes according to the state-transition relation:

$$x_{t+1} = f_t(x_t, u_t) \qquad \left(t = 0, 1, \ldots, T-1\right) \tag{2.39}$$

where $f_t : X \times U \to X$ is a given function, and X and U are the state space and control space, respectively. It is assumed that at each time period an objective function $g_t(x_t, u_t)$ is given that is real valued, and we are looking for a sequence of controls $(u_0, u_1, \ldots, u_T)$ such that an overall objective:

$$\sum_{t=0}^{T} g_t(x_t, u_t) \tag{2.40}$$

is optimal.

Without loss of generality, we may assume that function (2.40) is minimized; otherwise, we have to multiply it by –1. This problem can be solved as a nonlinear programming problem:

$$\begin{aligned} &\text{Minimize} \quad \sum_{t=0}^{T} g_t(x_t, u_t) \\ &\text{Subject to} \quad \left.\begin{aligned} x_t &\in X, \\ u_t &\in U \end{aligned}\right\} \quad t = 0,1,...,T \\ &\qquad\qquad x_{t+1} = f_t(x_t, u_t), \quad t = 0,1,...,T-1 \end{aligned} \tag{2.41}$$

In most applications, this problem has many decision variables and constraints, so its direct solution presents some computational difficulties. Instead of solving one large-scale problem, the *dynamic programming* (DP) procedure solves many smaller sized problems. The most popular DP method computes a sequence of optimal return functions $V_T(x)$, $V_{T-1}(x)$, …, $V_0(x)$ using the recursive equation:

$$V_t(x) = \min_{u \in U} \left\{ f_t(x,u) + V_{t+1}(f_t(x,u)) \right\} \tag{2.42}$$

with the initial term $V_{T+1}(x) = 0$ for all $x \in X$. Let $u(x)$ denote the optimizer of the right-hand side of Eq. (2.42), and let x_0 denote the initial state. Then, it can be proved that the sequence:

$$\begin{aligned} x_0^* &= x_0 \\ u_t^* &= u_t(x_t^*) \\ x_{t+1}^* &= f_t(x_t^*, u_t^*) \end{aligned} \tag{2.43}$$

gives a global optimum of the objective function, Eq. (2.40).

Example 2.18

Assume $x_0 = 0$; the state transition relation is:

$$\begin{aligned} &x_{t+1} = x_t + u_t, \qquad t = 0,1,...,T, \\ &X = U = R, \quad \text{and} \quad g_t(x_t, u_t) = (x_t - u_t + 1)^2 + x_t^2 \end{aligned}$$

In this case:

$$V_T(x) = \min_u g_T(x,u) = \min_u (x - u + 1)^2 + x^2$$

showing that:

$$u_T(x) = x + 1, \text{ and } V_T(x) = x^2$$

Then,

$$V_{T-1}(x) = \min_u \left\{ \left((x-u+1)^2 + x^2 \right) + (x+u)^2 \right\}$$
$$= \min_u \left\{ 3x^2 + 2u^2 + 2x - 2u + 1 \right\} = 3x^2 + 2x + \frac{1}{2}$$

with

$$u_{T-1}(x) = \frac{1}{2},$$

Similarly,

$$V_{T-2}(x) = \min_u \left\{ \left((x-u+1)^2 + x^2 \right) + 3(x+u)^2 + 2(x+u) + \frac{1}{2} \right\}$$
$$= \min_u \left\{ 5x^2 + 4ux + 4u^2 + 4x + \frac{3}{2} \right\} = 4x^2 + 4x + \frac{3}{2}$$

with $U_{T-2}(x) = -x/2$. This recursive reasoning should continue until $u_0(x)$ is obtained, and then the optimal decision sequence is obtained by Eq. (2.43).

In most applications, the optimization problem shown in Eq. (2.42) cannot be solved analytically, so functions $V_t(x)$ and $u_t(x)$ can be determined only numerically. For such cases, *discrete dynamic programming* (DDP) is used, where the state and decision spaces are discretized: $X = \{x^{(i)}\}$ and $U = \{u^{(j)}\}$. In this discrete case, recursive Eq. (2.42) is replaced by its discrete counterpart:

$$\hat{V}_t(x^{(i)}) = \min_j \left\{ f_t(x^{(i)}, u^{(j)}) + \hat{V}_{t+1}(f_t(x^{(i)}, u^{(j)})) \right\} \tag{2.44}$$

with the initial term $\hat{V}_{T+1}(x^{(i)}) = 0$ for all i. The difficulty in applying this procedure is that $f_t(x^{(i)}, u^{(j)})$, which represents the new state, does not need to be among the discrete nodes. This problem can be overcome by using an interpolation technique between the nodes. The main problem of using this method is known as the *curse of dimensionality*. As an illustration, assume that $T = 10$, and the state space is discretized into 10 levels in each of its 10 components. Then, the total number of nodes is 10^{10}. Therefore, DDP is used for only up to four or five decision and state variables.

In order to overcome the problem of dimensionality, successive approximation algorithms were developed, the most highly regarded techniques being *differential dynamic programming* (DIFF DP), *discrete differential dynamic programming* (DDDP), and *state incremental dynamic programming* (IDP). When applying any of these methods, an initial approximation of the optimal policy must be given by the user, and this approximate solution is improved in each step until convergence

occurs. However, as is usually the case in nonconvex optimization, the limit can be only a local optimum or the approximating sequence is divergent.

For the sake of simplicity let $\{x_t\}$ and $\{u_t\}$ denote the sequence of states and decisions of any iteration step and let $\{x_t^{new}\}$, $\{u_t^{new}\}$ be the successor obtained by applying an iteration step. In the case of DDDP, a discrete dynamic programming algorithm is used with the additional condition that for all t,

$$\left\| x_t - x_t^{new} \right\|_\infty < \varepsilon \tag{2.45}$$

where $\varepsilon > 0$ is a specified tolerance level, and $\|..\|_\infty$ is the maximum norm of the vectors. The additional constraints vastly reduce the number of discrete decision variable values which has to be tested in each step. However, iteration steps are repeated many times. Constraint (2.45) can also be described as the optimum of problem (2.44), which is determined only in a corridor of width ε centered about the predecessor. When applying IDP, we require that for all t,

$$x_t^{new} = x_t + \alpha_t e_i \tag{2.46}$$

where e_i is the ith basis vector, so x_t^{new} differs from x_t only in its ith coordinate. This additional constraint reduces the solution of optimization problem (2.42) to a one-dimensional line search. This procedure is very effective if the state transition relation is invertible in u_t; that is, there is a function h_t: $X \times X \mapsto U$ such that Eq. (2.39) is equivalent to the following equation:

$$h_t(x_t, x_{t+1}) = u_t \tag{2.47}$$

Then, Eq. (2.42) has the special form:

$$\hat{V}_t(\alpha_t) = \min_{\alpha_{t+1}} \left\{ g_t\left(x_t + \alpha_t e_i, h_t(x_t + \alpha_t e_i, x_{t+1} + \alpha_{t+1} e_j)\right) + \hat{V}_{t+1}(\alpha_{t+1}) \right\} \tag{2.48}$$

where only α_{t+1} is unknown. This one-dimensional optimization problem can be solved by any line-search algorithm.

We now take a look at the DIFF DP procedure by introducing the following notation:

$$\begin{aligned} \delta x_t &= x_t^{new} - x_t \\ \delta u_t &= u_t^{new} - u_t \end{aligned} \tag{2.49}$$

Let the quadratic Taylor's polynomial approximation of $f_T(x_t,u_t)$ be:

$$Q_T(x_T,u_T)=\frac{1}{2}D_T\delta x_T^2+E_T\delta x_T\delta u_T+\frac{1}{2}F_T\delta u_T^2+G_T\delta x_T+H_T\delta u_T+K_T \tag{2.50}$$

where for the sake of simplicity we assume that X and U are one-dimensional sets. Q_T is minimal if the first-order condition holds:

$$E_T\delta x_T+F_T\delta u_T+H_T=0$$

implying that:

$$\delta u_T=-\frac{H_T}{F_T}-\frac{E_T}{F_T}\delta x_T \tag{2.51}$$

The approximate optimal value function $\hat{V}_T(x_T)$ is now obtained as:

$$\hat{V}_T(x_T)=Q_T(x_T,u_T(x_T))=\frac{1}{2}(D_T-\frac{E_T^2}{F_T})(\delta x_T)^2+(G_T-\frac{E_TH_T}{F_T})(\delta x_T)$$

$$=\frac{1}{2}A_T(\delta x_T)^2+B_T(\delta x_T)$$

For $t < T$, the general step is similar to the previous construct. Assume that the approximate optimal return function:

$$\hat{V}_{t+1}(x_{t+1})=\frac{1}{2}A_{t+1}(\delta x_{t+1})^2+B_{t+1}(\delta x_{t+1}) \tag{2.52}$$

is already determined. Let $Q_t(x,u)$ be the quadratic Taylor's polynomial approximation of $g_t(x,u)+\hat{V}_{t+1}(f_t(x,u))$ about (x_t,u_t):

$$Q_t(x_t,u_t)=\frac{1}{2}D_t\delta x_t^2+E_t\delta x_t\delta u_t+\frac{1}{2}F_t\delta u_t^2+G_t\delta x_t+H_t\delta u_t+K_t \tag{2.53}$$

The optimum of this function can be obtained as it is presented in Eq. (2.50), with T being replaced by t. After this step is continued until $t = 0$, the successor policy can be obtained as:

$$x_0^{new}=x_0,\quad \delta x_0=0$$

and then

$$\delta u_t = -\frac{H_t}{F_t} - \frac{E_t}{F_t}\delta x_t$$

$$u_t^{new} = u_t + \delta u_t$$

$$x_{t+1}^{new} = f_t(x_t^{new}, u_t^{new})$$

for $t = 0, 1, 2, \ldots, T$. It can be proved that the DIFF DP method has quadratic convergence:

$$\left\| u^{new} - u^* \right\| < C \left\| u - u^* \right\|^2$$

where u^* is the optimal policy and $C > 0$ is a constant.

The survey paper by Yakowitz (1982) gives a summary of the different versions of dynamic programming with applications to water resources planning and management.

2.6.2 Genetic Algorithms

Genetic algorithms are very often used to find optimal solutions to computationally difficult problems. The basis of these methods is the similarity between biological and computational processes, especially in regard to how biological reproductive processes work. Genes and computers are similar to each other in the sense that they can record, copy, and disperse information. Just as in biological evolution, genetic algorithms work iteratively; many generations are created successively and only the fittest solutions survive. Using biological terminology, within the breeding population individual solutions are called *chromosomes*, which are encoded as strings (usually binary string). Individual features found in each chromosome are *genes*, and the value of a feature in any given chromosome is the *allele*.

In *creating* chromosomes we have to map the decision space into a set of encoded strings. The encoding mechanism depends on the problem being solved. One possible way is to use the binary representation of the decision variables. We also have to decide on the size of the breeding pool; that is, on the number of chromosomes being involved in the process. A larger number of chromosomes increases diversity, but the algorithm becomes slower. An initial population is first selected randomly.

The *fitness* of each string is often evaluated based on the objective functions, the constraints, and the encoding mechanism. A usual way is to add a penalty term to the objective function if any chromosome becomes not feasible by violating one or more constraints.

Pairs of chromosomes are formed randomly, and they are then subjected to certain genetic manipulations that modify the genes in the parent chromosomes. A typical procedure is *crossover*, which swaps a part of the genetic information contained in the two chromosomes. Usually a substring portion is selected randomly in

the chromosomes, and the genes (which are the string elements — for example, bits) within that substring are exchanged. In this way, new offsprings are created to replace the parent chromosomes. The particular structure of the crossover mechanism is usually problem specific and produces meaningful and feasible solutions.

There is no guarantee that the new pairs of chromosomes will be better than the parent chromosomes. To overcome this difficulty, *mutations* are allowed to occur which reverse one or more genes in a chromosome. This simple step may reintroduce certain genes into the solution that may be essential for the optimal solution and were lost from the breeding population in previous stages. Frequent use of mutations makes the search more broadly random, resulting in slower convergence of the algorithm.

By repeating the above steps iteratively, we sequentially replace either the entire previous population or only the less fit members of it. The cycle of creation, evaluation, selection, and manipulation is repeated iteratively until an optimal or an acceptable solution is reached.

Example 2.19

Assume function $f(x) = x - x^2$ is maximized in the [0,1] interval. First, the decision space is discretized and each discretized alternative is encoded. For illustration purposes, we have four-bit representations. The population is given in Table 2.7. The initial population is selected next. We have chosen six chromosomes as shown in Table 2.8, which also presents their actual values and the corresponding objective function values.

The worst objective function values are obtained at 0001 and 1111. A simple crossover procedure is performed by interchanging the two middle bits to obtain 0111 and 1001. The new population is given in Table 2.9. The worst objective function value is obtained at 0011, but two chromosomes have the second worst objective values: 1100 and 0100. Randomly selecting the first, we have the chromosome pair 0011 and 1100. By interchanging their middle bits we get 0101 and 1010. The resulting new population is shown in Table 2.10.

Selecting the pair 0100 and 0101, we cannot use the same crossover procedure as before, as they have the same middle bits. Because the two chromosomes differ in only one bit, interchanging any substring of them will result in the same chro-

TABLE 2.7
Population of Chromosomes

0000	0110
0001	1010
0010	1100
0100	0111
1000	1011
0011	1101
0101	1110
1001	1111

TABLE 2.8
Initial Population

Chromosomes	Value	Objective Function
0001	0.0625	0.0586
1100	0.75	0.1875
1011	0.6875	0.2148
1111	0.9375	0.0586
0100	0.25	0.1875
0011	0.1875	0.1523

TABLE 2.9
First Modified Population

Chromosomes	Value	Objective Function
0111	0.4375	0.2461
1100	0.75	0.1875
1011	0.6875	0.2148
1001	0.5625	0.2461
0100	0.25	0.1875
0011	0.1875	0.1523

TABLE 2.10
Second Modified Population

Chromosomes	Value	Objective Function
0111	0.4375	0.2461
0101	0.3125	0.2148
1011	0.6875	0.2148
1001	0.5625	0.2461
0100	0.25	0.1875
1010	0.625	0.2344

mosomes; therefore, mutation will be applied to them. By reversing the first two bits, we have 1000 and 1001, but the second chromosome is already in the population. Therefore, mutation is again used to reverse the second bit to obtain 1100. The resulting modified population is shown in Table 2.11.

It is interesting to see how the best objective function is evolving from population to population. Starting from the initial table with the value 0.2148, the value increased to 0.2461 and then remained the same. In Table 2.11 it again increased to 0.25. Because this value is the global optimum, it will not increase in any further

TABLE 2.11
Third Modified Population

Chromosomes	Value	Objective Function
0111	0.4375	0.2461
1100	0.75	0.1875
1011	0.6875	0.2148
1001	0.5625	0.2461
1000	0.5	0.25
1010	0.625	0.2344

steps. In practical applications, we stop the procedure if no or very small improvement occurs after a certain (user-specified) number of iterations.

More details on genetic algorithms and possible applications can be found in Goldberg (1989).

2.7 PROBLEMS

2.1 Graphically solve the following linear programming problem:

$$\begin{aligned} &\text{Minimize} && 2x_1 + x_2 \\ &\text{Subject to} && x_1, x_2 \geq 0 \\ & && x_1 + 3x_2 \geq 6 \\ & && 3x_1 + x_2 \geq 6 \end{aligned}$$

2.2 Consider the same feasible decision space as in the previous problem and minimize the nonlinear objective function:

(a) $x_1^2 + x_2^2$

(b) $x_1 x_2$.

2.3 Find the primal form for the following linear programming problem:

$$\begin{aligned} &\text{Minimize} && x_1 + 2x_2 - x_3 \\ &\text{Subject to} && x_1, x_3 \geq 3 \\ & && x_1 - x_2 + 2x_3 \geq 2 \\ & && x_1 + x_2 = 3 \\ & && 2x_1 - x_3 \leq 5 \end{aligned}$$

2.4 Consider the following problem with two objective functions:

$$\begin{aligned} &\text{Maximize} && f_1 = x_1 + x_2,\ f_2 = x_1 - x_2 \\ &\text{Subject to} && x_1, x_2 \geq 0 \\ & && x_1 + 3x_2 \leq 6 \\ & && 2x_1 - x_2 \geq 1 \end{aligned}$$

(a) Graphically show the decision space.
(b) Graphically show the criteria space.

2.5 Solve the previous problem with sequential optimization if f_1 is more important than f_2.
2.6 Select f_1 as the more important objective in problem 2.4. Give the range of ε_2 that leads to a feasible solution. Solve the problem for $\varepsilon_2 = 0$.
2.7 Solve problem 2.4 by using the weighing method, with $c_1 = c_2 = 0.5$. What are the solutions with and without normalizing the objectives?
2.8 Select equal weights in problem 2.4. Solve it by minimizing distance from the ideal point. What are your solutions with the ρ_1, ρ_2, and ρ_∞ distances?
2.9 Select $\underline{y} = (1,1)$ and $\underline{f}_* = (2,0)$, and apply the direction-based method to problem 2.4.
2.10 Consider the following problem with 2 objectives and 4 alternatives.

Alternatives	Objective 1	Objective 2
1	0	3
2	1	2
3	2	1
4	3	0

Assume both objectives are maximized and $c_1 = c_2 = 0.5$. Solve the problem with the application of:
(a) Weighting method
(b) Maximizing distance from the nadir (0,0) and selecting the ρ_1 distance.
(c) Repeat part (b) with the ρ_2 distance.
(d) Repeat part (b) with the ρ_∞ distance.

2.11 Consider the following social choice problem:

	Alternatives				
Decision Makers	**1**	**2**	**3**	**4**	**Weights**
1	1	3	2	4	0.3
2	4	1	2	3	0.2
3	4	3	2	1	0.2
4	1	4	3	2	0.2
5	3	2	1	4	0.1

Find the solution based on plurality voting.

2.12 Solve problem 2.11 using Borda counts.

2.13 Solve problem 2.11 using the Hare system.

2.14 Solve problem 2.11 using pair-wise comparisons, where the order of comparisons is 1, 2, 3, 4.

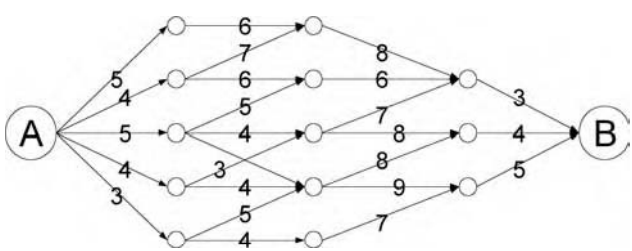
2.15 Assume that $\underline{d} = (0,0)$ and H is characterized by the following set of inequalities:

$$x_1, x_2 \geq 0$$
$$2x_1 + x_2 \leq 4$$
$$x_1 + 3x_2 \leq 6$$

Find the Nash solution, Eq. (2.37), of this problem.

2.16 Assume $c_1 = 1/3$ and $c_2 = 2/3$. Find the nonsymmetric Nash solution, Eq. (2.38), of the previous problem.

2.17 The following directed graph shows the possible arcs for a person moving from one point to another and the time periods required to walk through each arc. Assume somebody starts at point A and wants to reach point B in the shortest time period. Find the best path that minimizes the total time.



PROBLEM 2.17

2.18 Assume that for $t = 1, 2, 3, 4$, $g_t(x_t, u_t) = x_t^2 + u_t$, $x_t \in [0,1]$, $u_t \in [-1,1]$, and for all t, $x_{t+1} = x_t + u_t + 1$. Solve Eq. (2.41) by dynamic programming.
2.19 Repeat Example 2.19 for minimizing $f(x) = x^2 + x + 1$ in interval $[-1,1]$.
2.20 Solve problem 2.19 for minimizing $x^3 - 99x + 6$ on the integer points of interval [0,100]. You may use the binary representation of integers.

References

Dantzig, G. B., Linear Programming under Uncertainty, *Manage. Sci.*, 1(3), 157–168, 1955.
Dantzig, G. B., *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ, 1963.
Forgo, F., Szep, J., and Szidarovszky, F., *Introduction to the Theory of Games*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.
Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
Hadley, A., *Nonlinear and Dynamic Programming*, Addison-Wesley, Reading, MA, 1964.
Luce, R. D. and Raiffa, H., *Games and Decisions*, Wiley, New York, 1967.
Mangasarian, O. L., *Nonlinear Programming*, McGraw-Hill, New York, 1962.
Rubinstein, R. Y., *Simulation and Monte Carlo Method*, Wiley, New York, 1981.
Szidarovszky, F. M. and Bahill, A. T., *Linear Systems Theory*, 2nd ed., CRC Press, Boca Raton, FL, 1992.
Szidarovszky, F., Gershon, M., and Duckstein, L., *Techniques of Multi-Objective Decision Making in Systems Management*, Elsevier, Amsterdam, 1986.
Taylor, A. D., *Mathematics and Politics*, Springer-Verlag, New York, 1995.
Vajda, S., *Probabilistic Programming*, Wiley, New York, 1972.
Yakowitz, S., Dynamic programming applications in water resources, *Water Resources Res.*, 18(4), 673–696, 1982.

3 Decision Making under Uncertainty

This chapter is devoted mainly to the elements of probabilistic and statistical methodology and their use in modeling risk and uncertainty. We have also included a special section on neural networks, which are the most popular tools when models are uncertain or if no physical model is available.

3.1 INTRODUCTION

In all fields of natural resources management, we face uncertainty arising from natural phenomena that cannot be predicted accurately. Any prediction is uncertain, and in the mathematical modeling of hydrological and water resources systems this uncertainty has to be taken into account. Of the two primary ways to model uncertainty, one method is based on considering all uncertain elements as random variables and on the use of probabilistic and statistical models. The other approach is based on the theory of fuzzy sets, where uncertainty is modeled by membership functions. This chapter provides an introduction to this very important methodology.

3.2 PROBABILISTIC METHODS

Random events and random outcomes are usually characterized by random variables, the properties of which are the subject of the next subsection.

3.2.1 Random Variables

In natural resources management, we often face quantities for which the values are random, thus they are called *random variables*. A random variable is *discrete* if its possible values form a finite or a countable set. A discrete random variable, then, is completely defined by listing its possible values and giving the occurring probabilities: $X = x_1, x_2, \ldots$, and for $k = 1,2, \ldots, p_k = P(x = x_k)$. The p_k probability values are all taken from the unit interval [0,1], and their sum equals unity:

$$0 \le p_k \le 1 \quad \text{for} \quad k = 1,2,..., \quad \text{and} \quad \sum_k p_k = 1 \tag{3.1}$$

Example 3.1

The *Bernoulli* variable has two possible values: 1 and 0, with probabilities $p_1 = P(X = 1) = p$ and $p_0 = P(X = 0) = q\ (= 1 - p)$. For any random event A we can assign a Bernoulli variable $X = 1$ if A occurs; otherwise, $X = 0$, and $p_1 = P(X = 1) = P(A) = p$ and $p_0 = P(X = 0) = 1 - p$.

Example 3.2

Assume an experiment is repeated n times, independent of each other, and in each case an event A has the same probability, p. Let X denote the number showing how many times A occurs among the n trials. The possible values of X are 0, 1 , 2, …, n, with probabilities:

$$p_k = P(X = k) = \binom{n}{k} p^k q^{n-k}$$

This random variable is called *binomial*.

If $X_1, X_2, \ldots, X_n$ are independent Bernoulli variables with identical parameter p, then it is easy to see that $X = X_1 + X_2 + \ldots + X_n$ is binomial with parameters n and p.

Example 3.3

Assume that an experiment is repeated independently, and an event A has the same p probability in each case. Let X denote the number of repeats until A first occurs. The possible values of X are 1, 2, 3, … (until infinity), and $p_k = P(X = k) = q^{k-1}p$, where $q = 1 - p$ as before. This variable is *geometric*.

Example 3.4

Consider again the model of the previous example, and let $r \geq 1$ be a given positive integer. Now, let X denote the number of trials until A occurs r times. Clearly, the case of $r = 1$ coincides with the geometric variable. The possible values of X are r, $r + 1$, $r + 2$, … (until infinity), and

$$p_k = P(X = k) = \binom{k-1}{r-1} q^{k-r} p^r$$

for $k = r, r + 1, r + 2, \ldots$. This random variable is *negative binomial*.

Example 3.5

Assume a box contains N items, S of which are defective ($S < N$). If we randomly choose n items from the box *without* replacement, and X denotes the number of

defective items in the selected sample, then X is *hypergeometric*. The possible values of X are 0, 1, 2, …, min{S;n}, and the occurring probabilities are:

$$p_k = P(X = k) = \frac{\binom{S}{k}\binom{N-S}{n-k}}{\binom{N}{n}}$$

If we modify the above example by assuming that sampling is done *with* replacement, then each time an item is taken from the box we have the same $p = S/N$ probability for that item to be defective. Therefore, we have the binomial distribution with parameters n and p.

Example 3.6

Assume that in a binomial variable $n \to \infty$, $p \to 0$ so that np is kept on a constant λ level. The limiting distribution is called *Poisson.* The possible values of X are 0, 1, 2, … (until infinity) with probabilities:

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

Let A be any subset of the possible values of a discrete random variable X. Then:

$$P(X \in A) = \sum_{\{k | x_k \in A\}} P(X = x_k) \tag{3.2}$$

where the summation is done for all values of k such that $x_k \in A$. For example, if the possible values of X are integers, then:

$$P(k_1 \le X \le k_2) = \sum_{k=k_1}^{k_2} P(X = k) \tag{3.3}$$

where we assume that both k_1 and k_2 belong to the range of X.

The distribution function of a random variable is defined for all real values x:

$$F(x) = P(X \le x) = \sum_{\{k | x_k \le x\}} P(X = x_k) \tag{3.4}$$

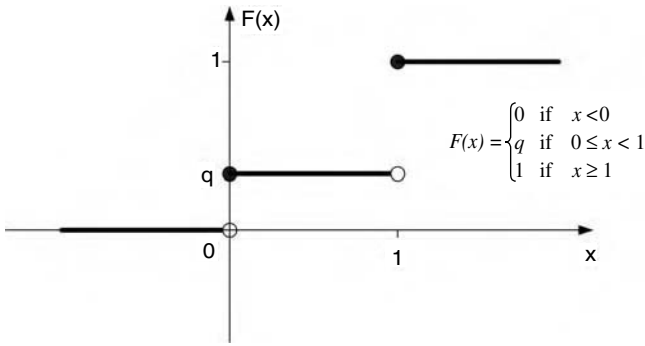


FIGURE 3.1 Distribution function of Bernoulli variables.

Clearly, if the possible values of X are integers, then

$$F(x) = \sum_{\{k|k\le x\}} P(X = k) \tag{3.5}$$

The graph of $F(x)$ for any discrete random variable is a piecewise constant function, starting with zero, and discontinuity occurs only at the values x_1, x_2, …, and they have a jump of height $p_k = P(X = x_k)$ at each point x_k. Figure 3.1 shows $F(x)$ for a Bernoulli variable.

Continuous random variables are defined directly by their distribution functions $F(x)$, which are assumed to be continuous at each x. Then, each particular value has zero occurring probability: $P(X = x) = 0$ for all x; furthermore,

$$P(a \le x \le b) = P(a < x \le b) = P(a \le x < b) = P(a < x < b) = F(b) - F(a). \tag{3.6}$$

The distribution functions of mixed distributions have discontinuities and are not piecewise constant functions. They are not very common in water resources applications.

The density function of a continuous variable is defined as the derivative of $F(x)$:

$$f(x) = F'(x) \tag{3.7}$$

which exists except as a zero-measured set. By the fundamental theorem of integral calculus we see that

$$F(b) - F(a) = \int_a^b f(x)dx$$

so probabilities given in Eq. (3.6) can be also obtained by integrating the density function.

The distribution function and density function of any continuous variable satisfy the following properties:

$$\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to+\infty} F(x) = 1, \quad \text{and}$$

$$f(x) = F'(x) \geq 0, \text{ furthermore } \int_{-\infty}^{\infty} f(x)dx = 1$$

Example 3.7

A random variable is *uniform* in a finite interval [*a*,*b*] if its distribution function is:

$$F(x) = \begin{cases} 0 & if \quad x \leq a \\ \dfrac{x-a}{b-a} & if \quad a < x \leq b \\ 1 & if \quad x > b \end{cases} \tag{3.8}$$

By differentiation,

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if} \quad a < x < b \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

If $\alpha,\beta \in [a,b]$, then:

$$P(\alpha \leq X \leq \beta) = F(\beta) - F(\alpha) = \frac{\beta-a}{b-a} - \frac{\alpha-a}{b-a} = \frac{\beta-\alpha}{b-a} \tag{3.10}$$

showing that this probability depends only on the length $\beta - \alpha$ of the interval $[\alpha,\beta]$ and is independent of the location of this interval. This is the reason for calling this variable *uniform*.

Example 3.8

A continuous random variable is called *exponential* if:

$$F(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ 1 - e^{-\lambda x} & \text{if} \quad x > 0 \end{cases} \tag{3.11}$$

where $\lambda > 0$ is a given parameter. It is easy to see that

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if} \quad x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.12}$$

The exponential distribution has the forgetfulness property, which means the following: If $0 < x < x + y$, then

$$P(X > x + y | X > y) = P(X > x) \tag{3.13}$$

showing that if X, for example, represents the lifetime of any entity, then at any age y surviving an additional x time periods does not depend on the age.

Because of the forgetfulness property, exponential variables are only very seldom used in lifetime modeling. For such purposes, a gamma or a Weibull distribution is used, which are introduced next.

Example 3.9

Let $\alpha, \lambda > 0$ be two given parameters. The gamma variable that has these parameters is defined by the density function:

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & \text{if} \quad x > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.14}$$

where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

is the gamma function. Let $\alpha = n$ be an integer and $X_1, X_2, \ldots, X_n$ be independent exponential variables with the same parameter λ. Then, $X_1 + X_2 + \ldots + X_n$ follows a gamma distribution with parameters $\alpha = n$ and λ.

The distribution functions of gamma variables cannot be given in closed-form, simple expressions except in very special cases. They have been tabulated, so function tables are available. Alternatively, numerical integration can be used.

Example 3.10

Let $\lambda, \beta > 0$ be given parameters. The Weibull distribution with these parameters is defined by the distribution function:

$$F(x) = \begin{cases} 1 - e^{-\lambda x^{\beta}} & \text{if} \quad x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

By differentiation,

$$f(x) = \begin{cases} \lambda\beta x^{\beta-1} e^{-\lambda x^{\beta}} & \text{if} \quad x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.16}$$

It is easy to see that, if Y is exponential with parameter λ, then $X = Y^{1/\beta}$ is Weibull with parameters λ and β.

Example 3.11

The *standard normal* variable is given by the density function:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad (-\infty < x < \infty) \tag{3.17}$$

The distribution function cannot be given by an analytic form. It is tabulated and is usually denoted by $\phi(x)$. A general *normal* variable is obtained as follows. Let μ and $\sigma > 0$ be two given parameters and Z be a standard normal variable. Then, $X = \sigma Z + \mu$ follows a normal distribution with parameters μ and σ. It is easy to see that, in general,

$$F(x) = \phi\left(\frac{x - \mu}{\sigma}\right) \tag{3.18}$$

and

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.19}$$

Figure 3.2 shows the density and distribution function of the most frequently used distributions.

In statistical methods, we often use distributions arising from the normal variable. The three most important such variables are given next.

Example 3.12

If $Z_1, Z_2, \ldots, Z_n$ are independent standard normal variables, then:

$$X_n = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \tag{3.20}$$

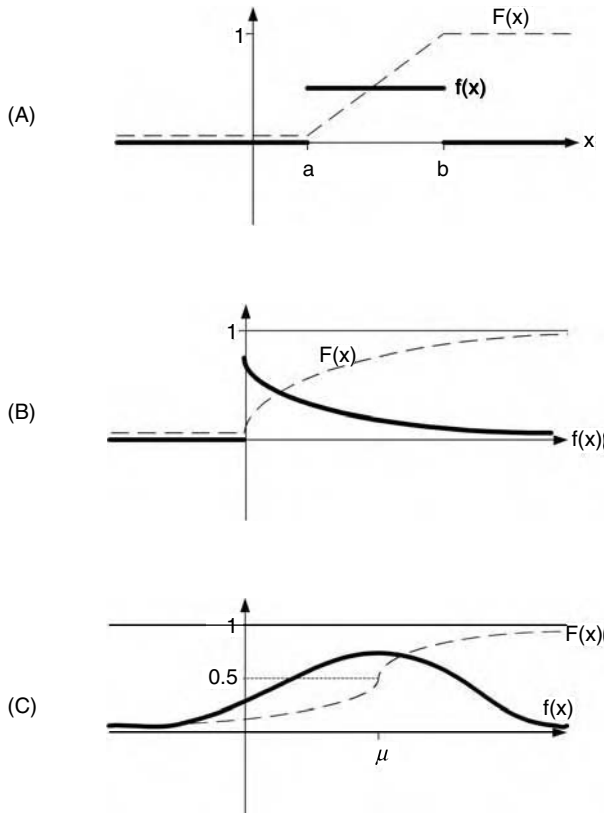


FIGURE 3.2 (A) Uniform distribution; (B) exponential distribution; (C) normal distribution.

is the *chi-square* distribution with n degrees of freedom. The corresponding density function is the following:

$$f(x) = \begin{cases} \dfrac{\frac{1}{2} e^{\frac{-x}{2}} \left(\frac{x}{2}\right)^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)} & \text{if} \quad x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.21}$$

By comparing this density to the gamma density given in Eq. (3.14) it is clear that chi-square is a special gamma distribution with $\alpha = n/2$ and $\lambda = 1/2$.

Example 3.13

Let Z be a standard normal variable and X a chi-square variable with n degrees of freedom. Then,

$$T_n = \frac{Z}{\sqrt{X/n}} \tag{3.22}$$

is the *t-distribution* with n degrees of freedom. It can be shown that the density function of T_n is:

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad (-\infty < 0 < \infty) \tag{3.23}$$

The shape of $f(x)$ is very similar to that of the standard normal density function.

Example 3.14

Let X_n and X_m be independent chi-square variables with n and m degrees of freedom, respectively. Then,

$$F_{n,m} = \frac{X_n/n}{X_m/m} \tag{3.24}$$

is the *F-distribution* with n and m degrees of freedom. The density function of $F_{n,m}$ is as follows:

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\frac{n+m}{2}\right) n^{n/2} m^{m/2} x^{(n-2)/2}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)(m+nx)^{\frac{n+m}{2}}} & x > 0 \\ 0 \qquad \text{otherwise.} \end{cases} \tag{3.25}$$

The distribution functions of the above three distributions cannot be presented in simple equations so they have been tabulated and their particular values can be found in the tables. For more details and tables, see Milton and Arnold (1995) or Ross (1987).

3.2.2 Central Tendencies and Variances

Consider first a discrete random variable X with values $x_1 < x_2 < x_3 < \ldots$ and occurring probabilities $p_k = p(X = x_k)$. Central tendencies are characterized by expectation, median, and mode. The *expectation* of X is defined as:

$$E(x) = \sum_k x_k p_k \tag{3.26}$$

by assuming that

$$\sum_k |x_k| p_k < \infty$$

That is, $E(x)$ is a weighted average of the possible values of X where the weights are the occurring probabilities.

The *median* of X is a particular value x_l such that:

$$p_1 + p_2 + \cdots + p_{l-1} < \frac{1}{2} \le p_1 + p_2 + \cdots + p_{l-1} + p_l \tag{3.27}$$

That is, the median is the x_l value for which

$$P(X \le x_{l-1}) < \frac{1}{2} \le P(X \le x_l)$$

or where the cumulative probability first reaches or exceeds 1/2.

The *mode* of X is the value x_m such that

$$p_m = \max_k \{p_k\} \tag{3.28}$$

which is the value of X that has the highest probability of occurring. If more than one value has the same largest probability, then the mode consists of all such values.

Now let g be any real valued function. Then, more generally,

$$E\big(g(x)\big) = \sum_k g(x_k) p_k \tag{3.29}$$

assuming that

$$\sum_k |g(x_k)| p_k < \infty$$

As special cases, $E(X)$, $E(X^2)$, $E(X^3)$, … are called the first, second, third, … *moments* of X. The real valued function:

$$m(t) = E\big(e^{tX}\big) = \sum_k e^{tx_k} p_k \tag{3.30}$$

is the *moment-generating function* of X. It is easy to see that

$$\begin{aligned} E(X) &= m'(0) \\ E(X^2) &= m''(0) \\ E(X^3) &= m'''(0) \end{aligned} \tag{3.31}$$

and so on; therefore, the moment-generating function is very useful in computing the moments.

In the case of a continuous random variable X, let f denote the density function. Then, the *expectation* of X is given as:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{3.32}$$

and, more generally, for any integrable function g,

$$E(g(x)) = \int_{-\infty}^{\infty} g(x)f(x)dx \tag{3.33}$$

The *median* of X is the value x such that

$$F(x) = \frac{1}{2} \tag{3.34}$$

If F has a constant segment at the 1/2 level, then the smallest x value satisfying this equation is chosen. The *mode* of X is the maximizer of its density function; that is, the mode of X is x^* if:

$$f(x^*) = \max_x \{f(x)\} \tag{3.35}$$

If the maximizer is not unique, then all of them form the mode. The definition of the moment-generating function is analogous to Eq. (3.30):

$$m(t) = E\left(e^{tX}\right) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \tag{3.36}$$

and because Eq. (3.31) remains true in the continuous case, it can be easily used to compute the moments of the distribution.

The variability of a random variable is characterized by its *variance*. If we use the notation $\mu = E(X)$, then the variance of X is defined to be:

$$Var(X) = E\left[(X-\mu)^2\right] \tag{3.37}$$

It can be shown that $Var(X)$ can be easily computed from the first two moments:

$$Var(\mathrm{X}) = E\left(X^2\right) - \mu^2 \tag{3.38}$$

In Table 3.1 we summarize the expectations and variances of the most popular distribution types. More details can be found, for example, in Ross (1987).

TABLE 3.1
Expectations and Variances

Name	$E(x)$	$Var(x)$
Bernoulli	p	pq
Binomial	np	npq
Geometric	$\frac{1}{p}$	$\frac{q}{p^2}$
Negative binomial	$\frac{r}{p}$	$\frac{rq}{p^2}$
Hypergeometric	$n\frac{s}{N}$	$n\frac{s}{N}\left(1-\frac{s}{N}\right)\left[1-\frac{n-1}{N-1}\right]$
Poisson	Λ	λ
Uniform	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Weibull	$\lambda^{-\frac{1}{\beta}}\Gamma\left(1+\frac{1}{\beta}\right)$	$\lambda^{-\frac{2}{\beta}}\Gamma\left(1+\frac{2}{\beta}\right)-\mu^2$
Normal	μ	σ^2
Chi-square	n	$2n$
t	0 (if $n>1$)	$\frac{n}{n-2}$ (if $n>2$)
F	$\frac{m}{m-2}$ (if $m>2$)	$\frac{m^2(2m+2n-4)}{m(m-2)^2(m-4)}$ (if $m>4$)

3.2.3 Joint Distributions

Let X and Y be two *discrete* distributions with possible values $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$, respectively. Their joint distribution is defined by the joint probabilities:

$$p_{ij} = P\Big((X = x_i) \cap (Y = y_j)\Big) \tag{3.39}$$

Similar to the single-variables case, these p_{ij} values satisfy the following relations:

$$\begin{aligned} &0 \le p_{ij} \le 1 \quad \text{for all } i \text{ and } j, \\ &\sum_i \sum_j p_{ij} = 1. \end{aligned} \tag{3.40}$$

If $A \subseteq R^2$ is any set, then the probability that the pair (X,Y) belongs to A is given by the sum:

$$P\big((X,Y) \in A\big) = \sum_{\big\{(i,j)\big|(x_i,y_j)\in A\big\}} p_{ij} \tag{3.41}$$

The *marginal* probabilities of X and Y can be easily computed as:

$$p_i^x = P(X = x_i) = \sum_j p_{ij} \tag{3.42}$$

and

$$p_j^Y = P(Y = y_j) = \sum_i p_{ij} \tag{3.43}$$

Variables X and Y are *independent* if, for all i and j,

$$p_{ij} = p_i^x \cdot p_j^y \tag{3.44}$$

If g is any bivariable function, then:

$$E\big[g(X,Y)\big] = \sum_i \sum_j g(x_i, y_j) p_{ij} \tag{3.45}$$

where we assume that

$$\sum_i \sum_j \left| g\left(x_i, y_j\right) \right| p_{ij} < \infty$$

similar to the single-variable case.

Continuous joints distributions are characterized by a joint density function f, which satisfies the following properties:

$$f(x, y) \geq 0 \text{ for all } x \text{ and } y,$$

and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1$$

Again, let $A \subseteq R^2$ be any set. Then, similar to Eq. (3.41),

$$P\left((X, Y) \in A\right) = \iint_A f(x, y) dy dx \tag{3.46}$$

The marginal density functions of X and Y are obtained as:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \tag{3.47}$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \tag{3.48}$$

Variables X and Y are *independent* if, for all x and y,

$$f(x, y) = f_x(x) f_y(y) \tag{3.49}$$

Let g be any integrable bivariable function. Then,

$$E\left[g(X, Y)\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dy dx \tag{3.50}$$

Expectations and variances satisfy the following important properties:

1. $E(\alpha X + \beta) = \alpha E(X) + \beta$ for all real α and β.
2. $E(X + Y) = E(X) + E(Y)$.
3. If X and Y are independent, then $E(XY) = E(X)\mathrm{E}(Y)$.
4. $Var(\alpha X + \beta) = \alpha^2 \mathrm{Var}(X)$ for all real α and β.
5. If X and Y are independent, then $Var(X + Y) = Var(X) + Var(Y)$.

Note that, without independence of X and Y, properties (3) and (5) might not be true.

The *covariance* of X and Y is defined as:

$$Cov(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) \tag{3.51}$$

Clearly, if X and Y are independent, then property (3) implies that $Cov(X,Y) = 0$.

The *correlation* between X and Y is determined in the following way:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \tag{3.52}$$

If X and Y are independent, then $Corr(X,Y) = 0$, and always $-1 \le Corr(X,Y) \le 1$.

It is easy to show that if a deterministic, linear functional relation exists between X and Y such that with some constants $\alpha \neq 0$ and β, $Y = \alpha X + \beta$, then:

$$Corr(X,Y) = \begin{cases} 1 & \text{if} \quad \alpha > 0 \\ -1 & \text{if} \quad \alpha < 0 \end{cases}$$

Based on the above properties, correlation is usually used to measure the strength of the relation between two random variables. If $|\mathrm{Corr}(X,Y)|$ is small, then the relation is very weak; if $Corr(X,Y)$ is close to +1 or –1, then we conclude that a strong relation exists between X and Y.

The *conditional distribution* of X given the value of Y is defined as the conditional probabilities

$$P\left(X = x_i \middle| Y = y_j\right) = \frac{p_{ij}}{p_j^Y} \tag{3.53}$$

in the discrete case, and as the conditional density function:

$$f_x\left(x|y\right) = \frac{f(x,y)}{f_Y(y)} \tag{3.54}$$

in the continuous case. The conditional distribution of Y given the values of X is defined analogously. With any fixed value of Y, we can obtain the *conditional*

expectation of X by using the conditional distribution given previously. In the discrete case,

$$E\left(X\middle|Y=y_j\right)=\frac{1}{p_j^Y}\sum_i x_i p_{ij} \tag{3.55}$$

assuming that

$$\sum_i |x_i| p_{ij} < \infty$$

and in the continuous case,

$$E(X|Y=y)=\frac{1}{f_Y(y)}\int_{-\infty}^{\infty} x f(x,y)dx \tag{3.56}$$

The conditional expectation of Y given the value of X can be obtained in a similar manner. These conditional expectations are usually referred to as *regression curves*.

We mention, finally, that Milton and Arnold (1995) and almost all introductory textbooks on probability theory offer extended discussions of joint distributions and their characteristics.

3.3 SIMULATION MODELS

When analyzing complex systems, the physical model and the mathematical description is so complicated that no solution algorithm can be developed. In many of these cases, no mathematical model is even available. The only choice we have then is simulation of the model. This section first introduces the most popular model simulation method, *neural networks*. After a model is formulated, its parameters are determined by observations, measurements, predictions, or computations. Such data usually are not available directly or are uncertain. In the first case, some kind of observer is used, and in the second case the usual method is to utilize probabilistic methods and simulation. Natural phenomena can very seldom be described accurately; for example, future floods, rainfall, and so on can be modeled only by probabilistic methods. The second part of this section will introduce the fundamentals of observers and stochastic simulation.

3.3.1 Neural Networks

Every mathematical model is an input–output relation in a certain sense, where the input is the set of model parameters and our actions and the output is the consequence of our actions. In optimization models, the best action (input) is selected based on

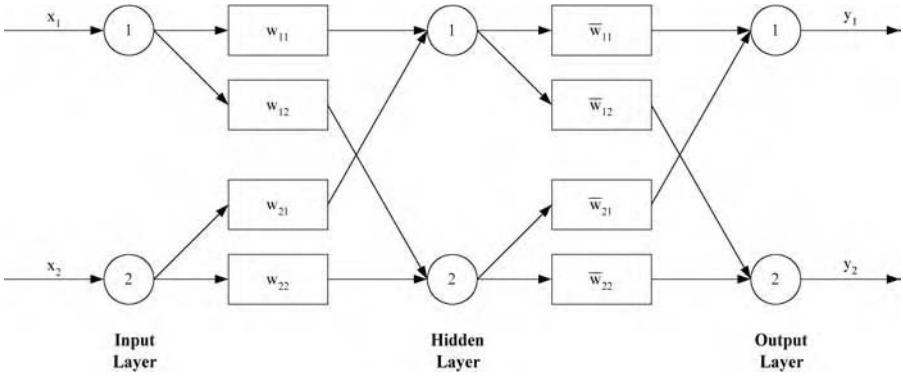


FIGURE 3.3 A simple neural network.

the given criteria. Input–output relations can be very conveniently modeled by the neural network methodology.

A simple neural network is shown in Figure 3.3, where the input variables are in the nodes of the input layer and the output variables are in the nodes of the output layer; we assume two input and output variables. In order to provide a high level of flexibility in model formulation, we also introduce hidden layers in the middle; these nodes correspond to hidden or artificial variables.

Let z_1 and z_2 denote the artificial variables corresponding to the hidden layer nodes. In the case of linear neural networks, we assume that for all hidden variables:

$$z_j = \sum_{i=1}^{m} w_{ij} x_i \tag{3.57}$$

where m is the number of input variables. In our case, $m = 2$. The coefficients w_{ij} are constants, which are determined by a training procedure to be explained later. The relations between the output and hidden variables are also linear:

$$y_k = \sum_{j=1}^{l} \overline{w}_{jk} z_j \tag{3.58}$$

where l is the number of hidden variables and we have n output variables. The coefficients $\overline{w}_{jk}$ are also estimated by the training process. For the nonlinear neural network, Eqs. (3.57) and (3.58) have certain nonlinear elements. The parameters of the nonlinear functions are also determined by the training process. In most applications, more than one hidden layer is introduced, and relations similar to Eqs. (3.57) and (3.58) are assumed to exist between the variables of the consecutive hidden layers. After the parameters of these function relations are estimated, it is very simple to compute the output $\underline{y} = (y_1, \ldots, y_n)$ for any input vector $\underline{x} = (x_1, \ldots, x_m)$. We can simply substitute the components of $\underline{x}$ into Eq. (3.57) to obtain the hidden variables,

and then these results are substituted into Eq. (3.58) to obtain the output variables. In the case of more hidden layers, additional similar steps are necessary. Because the input–output relation is easy to compute, the neural network model can be considered as a simple subroutine of input–output computations in any descriptive or optimization procedure.

The theoretical basis of the previously described neural network structure is the famous Kolmogorov's theorem (see, for example, Hecht–Nielsen, 1990), which states that any continuous function $f:[0,1]^m \mapsto R^n$ can be implemented exactly by a three-layer neural network with m input nodes, n output nodes, and $(2m + 1)$ hidden nodes, where Eqs. (3.57) and (3.58) have certain special (usually irrational) nonlinear elements. If these nonlinear functions are approximated by linear relations, then we need more terms; therefore, more hidden variables are required which might be structured in more than one hidden layer (Bose and Liang, 1996; Hecht-Nielsen, 1990).

As has been mentioned earlier, the application of neural network methodology depends on the estimation of the model parameters w_{ij} and $\overline{w}_{jk}$. The most frequently applied methodology is *training the network*, which is an iteration procedure that successively improves the parameter values. One possible method is the following. In the first step, we change the weights between the output layer and the hidden layer by using the following equation:

$$\overline{w}_{jk}(t+1) = \overline{w}_{jk}(t) + \beta E_k(t) y_k(t) \tag{3.59}$$

where β is the learning rate, $y_k(t)$ is the actual output, and $E_k(t) = d_k(t) - y_k(t)$ is the difference between the desired and actual outputs. The same equation cannot be used to estimate the weights between the input and hidden layers; the hidden variables are unknown, so we do not have their values. Instead, the following rule is used:

$$w_{ij}(t+1) = w_{ij}(t) + \beta z_j(t) \sum_k \overline{w}_{jk}(t) E_k(t) \tag{3.60}$$

This updating process is used iteratively for a large number of input–output pairs until the error terms $E_k(t)$ become sufficiently small.

3.3.2 Observers

In analyzing practical systems such as those arising in water resources management, some state variables cannot be directly measured. However, by constructing a special feedback structure the state can be made indirectly available. Such feedback structures are called *observers*. Let

$$\begin{aligned} \dot{\underline{x}} &= \underline{A}\underline{x} + \underline{B}\underline{u} \\ \underline{y} &= \underline{C}\underline{x} \end{aligned} \tag{3.61}$$

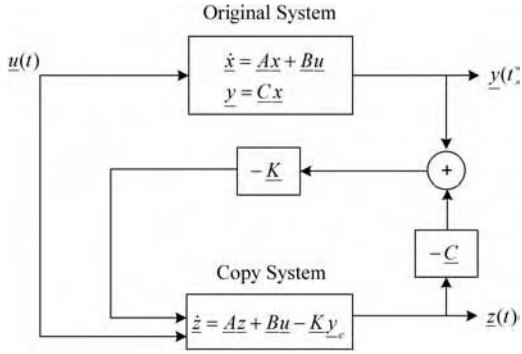


FIGURE 3.4 Feedback structure of an observer.

be a time-invariant continuous linear system. Assume that its input $\underline{u}(t)$ and output $\underline{y}(t)$ are available for any $t \geq 0$, but we have no way to measure its state variable, $\underline{x}(t)$. By building an identical system, we hope that its state will be a good estimate for the state of the original system; however, in the case of unstable systems this is not the case. Assume that the copy of the system is governed by the systems equation:

$$\dot{\underline{z}} = \underline{A}\underline{z} + \underline{B}\underline{u} \tag{3.62}$$

Subtract Eq. (3.62) from (3.61) to see that

$$\dot{\underline{x}} - \dot{\underline{z}} = \underline{A}(\underline{x} - \underline{z})$$

so the discrepancy between the states satisfies a homogeneous equation. It is well known from systems theory (see, for example, Szidarovszky and Bahill, 1992), that

$$\underline{x}(t) - \underline{z}(t) = e^{\underline{A}(t-t_0)}\left(\underline{x}(t_0) - \underline{z}(t_0)\right)$$

so $\underline{x}(t) - \underline{z}(t)$ may become very large for increasing values of t. This difficulty can be overcome by the feedback structure shown in Figure 3.4.

By subtracting the systems equations of the original and copy systems we obtain:

$$\dot{\underline{x}} - \dot{\underline{z}} = (\underline{A}\underline{x} + \underline{B}\underline{u}) - (\underline{A}\underline{z} + \underline{B}\underline{u} - \underline{K}\underline{y}_e) = \underline{A}(\underline{x} - \underline{z}) + \underline{K}(\underline{C}\underline{x} - \underline{C}\underline{z}) = (\underline{A} + \underline{K}\underline{C})(\underline{x} - \underline{z}) \tag{3.63}$$

If the original system is observable, then a constant matrix $\underline{K}$ exists such that all eigenvalues of $\underline{A} + \underline{K}\underline{C}$ have negative real parts implying that $\underline{x}(t) - \underline{z}(t) \to \underline{0}$ as $t \to \infty$. That is, for large values of t, the online measurements of the state of the copy system are the same as for the state of the original system, so we are able to observe $\underline{x}(t)$ by measuring the value of $\underline{z}(t)$. This theoretical basis for constructing

observers is known as the *eigenvalue placement theorem* (see, for example, Szidarovszky and Bahill, 1992, pp. 352–353).

3.3.3 Stochastic Simulation

The consequences of events affected by random elements are usually assessed by using simulation methods. The basis of such methods is the random number generators of computers which produce uniformly distributed, independent random numbers between 0 and 1. However, most random elements in nature have different distributions, so these random numbers cannot be used directly. A transformation method is usually used to obtain random values from other than uniform distributions.

Consider first a discrete random variable with values $x_1, x_2, \ldots$ and probabilities of occurring $p_1, p_2, \ldots$. Divide the [0,1] intervals into subintervals $[0, p_1)$, $[p_1, p_1 + p_2)$, $[p_1 + p_2, p_1 + p_2 + p_3)$, and so on, and let $I_1, I_2, \ldots$ denote these intervals. If U is a uniform variable in [0,1], then $P(U \in I_k) = p_k$.

Notice that $I_1 \cup I_2 \cup \ldots = [0,1)$, and if $U \neq 1$, then U belongs to exactly one of the subintervals. Define the random variable X as:

$$X = x_k \quad \text{if} \quad U \in I_k \tag{3.64}$$

This is the random variable we wish to generate, as its values are the x_k numbers, $k = 1, 2, \ldots$, and $P(X = x_k) = P(U \in I_k) = (p_1 + p_2 + \ldots + p_k) - (p_1 + p_2 + \ldots + p_{k-1}) = p_k$.

Example 3.15

Let X be a Bernoulli variable with probability p. That is, the value of X is either 1 or 0, and

$$P(X = 1) = p \quad \text{and} \quad P(X = 0) = q = (1 - p)$$

In this case $I_1 = [0,p)$ and $I_2 = [p,1)$. So, from a uniform random number, X can be obtained as:

$$X = \begin{cases} 1 & \text{if} \quad U < p \\ 0 & \text{otherwise.} \end{cases} \tag{3.65}$$

Example 3.16

Binomial variables can be generated by using the above general method. However, we can get the same result by using the following simple idea. Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli variables with the same parameter p. Then, $X = X_1 + X_2 + \ldots + X_n$ is a binomial variable with parameters n and p. Therefore, the simple procedure of the previous example is used n times, and the resulting random numbers are added.

Example 3.17

Poisson variables have only one parameter, λ, and take on values 0, 1, 2, … with probabilities $p_k = P(X = k) = (\lambda^k/k!)e^{-\lambda}$. One may proceed with the general method and taking advantage of the recursive relation $p_k = p_{k-1}\lambda/k$. However, exponential random variables can be effectively used here as well. The way to simulate exponential variables will be discussed later.

Let $X_1, X_2, \ldots$ be a sequence of independent exponential random variables with the same λ as a parameter, so $E(X_k) = 1/\lambda$ for all k, and add the X_k values until:

$$\sum_{i=1}^{k-1} X_i < 1 \le \sum_{i=1}^{k} X_i \tag{3.66}$$

Then $X = k$ gives a random value of the Poisson variable with the same parameter value λ.

Continuous random variables are generated by using their cumulative distribution functions. Let X be a random variable with distribution function $F(x)$, let U be uniform in [0,1], and let u be a random value of U. Then, the solution of equation $F(x) = u$ for x gives a random value of X. In order to see this, consider the distribution function of X at any real value t:

$$P(X \le t) = P\big(F(X) \le F(t)\big) = P\big(U \le F(t)\big) = F(t)$$

because U is uniform in [0,1]. Here, we assumed that F is strictly increasing; therefore, equation $F(x) = u$ has a unique solution for all $0 < u < 1$. The events $U = 0$ and $U = 1$ occur with zero probability.

Example 3.18

If X is exponential with parameter λ, then $F(x) = 1 - e^{-\lambda x}$, so we have to solve equation:

$$1 - e^{-\lambda x} = u$$

which implies that:

$$x = -\frac{1}{\lambda}\ln(1-u) \tag{3.67}$$

Example 3.19

Random values for a Weibull variable can be obtained simply by using exponential variables and the definition of a Weibull distribution. It is well known that if Y is

exponential with parameter α, then $X = Y^{1/\beta}$ with some $\beta > 0$ is a Weibull variable with parameters α and β. Then,

$$x = \left[-\frac{1}{\lambda} \ln(1-u) \right]^{1/\beta} \tag{3.68}$$

gives a random Weibull value, where u is a uniform value in [0,1].

Example 3.20

It is well known that if X_k (k = 1, 2, …, n) are independent exponential variables with the same λ parameter, then $X_1 + X_2 + \ldots + X_n$ is a gamma variable with parameters $\alpha = n$ and $\beta = 1/\lambda$. Hence, we must generate independent uniform variable values u_1, u_2, …, u_n, then:

$$X = \sum_{k=1}^{n} x_k = -\frac{1}{\lambda} \sum_{k=1}^{n} \ln(1-u_k) = -\frac{1}{\lambda} \ln\left[(1-u_1)(1-u_2) \cdots (1-u_n) \right] \tag{3.69}$$

follows the given gamma distribution.

Example 3.21

For a normal distribution, the general method cannot be used easily, as the standard normal distribution function Φ is not given in analytic form; it is only tabulated. Therefore, the solution of the equation:

$$\Phi\left(\frac{x-\mu}{\sigma} \right) = u$$

requires the use of a function table, which can be included in the general software. An easier approach is offered by the *central limit theorem*, which implies that if u_1, u_2, …, u_n are independent uniform numbers in [0,1], then for large value of n,

$$Z = \frac{\sum_{k=1}^{n} u_k - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \tag{3.70}$$

is a standard normal value. So if μ and σ^2 are given, then

$$X = \sigma Z + \mu \tag{3.71}$$

follows the normal distribution with mean μ and variance σ^2.

Consider now any process that depends on random elements. Let $X_1, X_2, \ldots, X_n$ denote the random variables of the process. Assume that the outcome is a function of these variables $Y = f(X_1, \ldots, X_n)$, where function f might be a given function or represents a computational procedure such as given by neural networks. For any random values $x_1^{(i)}, \cdots, x_n^{(i)}$ of $X_1, X_2, \ldots, X_n$, the corresponding values of Y can be obtained as $y^{(i)} = f(x_1^{(i)}, \cdots, x_n^{(i)})$. Let N be the number of simulations, so $i = 1, 2, \ldots, N$. The values $y^{(1)}, y^{(2)}, \ldots, y^{(N)}$ can be considered as a random sample of Y, so standard statistical methods can be used to examine the characteristics of the random outcome. In most cases, we are interested in $E(Y)$, which is estimated by the sample mean:

$$E(Y) \approx \frac{1}{N}\sum_{i=1}^{N} y^{(i)} = \bar{y} \tag{3.72}$$

so the sample mean is accepted as the expected value. The accuracy of this estimation can be characterized by its variance:

$$Var(\bar{y}) = \frac{1}{N} Var\left(f\left(X_1, \cdots, X_n\right)\right) \tag{3.73}$$

which converges to zero as $N \to \infty$. If N is sufficiently large, then $\bar{y}$ can be considered a normal variable as the result of the central limit theorem. Therefore, we have for any $\varepsilon > 0$,

$$P\left(\left|\bar{y} - E(Y)\right| < \varepsilon\right) = P\left(-\varepsilon < \bar{y} - E(Y) < \varepsilon\right) = P\left(-\frac{\varepsilon\sqrt{N}}{\sigma} < \frac{\bar{y} - E(Y)}{\sigma/\sqrt{N}} < \frac{\varepsilon\sqrt{N}}{\sigma}\right)$$

where $\sigma^2 = Var(f(X_1, \ldots, X_n))$. If Φ again denotes the standard normal distribution function, then this probability can be further simplified as:

$$\Phi\left(\frac{\varepsilon\sqrt{N}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon\sqrt{N}}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{N}}{\sigma}\right) - 1 \tag{3.74}$$

This is the probability that $\bar{y}$ approximates the true expectation $E(Y)$ within the error bound ε. Notice that as $N \to \infty$, this probability value converges to $2\Phi(\infty) - 1 = 1$.

Rubinstein (1981) gives a comprehensive summary of random number generators and simulation methods.

3.4 MEASURES OF RELIABILITY

Reliability theory is concerned with the statistical properties of the lifetimes of products, machinery, hydrological objects, systems, and so on. Because the lifetime

depends on the particular object or system we consider, it must be considered to be a random variable. Let X be the time until the first failure occurs. It is well known that the *distribution function* of X is defined as:

$$F(t) = P(X \le t), \tag{3.75}$$

which represents the probability that the object or system breaks down before time period t. The reliability function gives the probability that it will work without breaking down at least until time period t:

$$R(t) = P(X > t) = 1 - P(X \le t) = 1 - F(t) \tag{3.76}$$

In reliability theory, the Weibull distribution is assumed to model lifetime most appropriately. In this case, (see Example 3.10):

$$F(t) = 1 - e^{-\alpha t^{\beta}}, \ (t > 0) \tag{3.77}$$

where α, $\beta > 0$ are given parameters. The density function is obtained by simple differentiation:

$$f(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}} \ \ (t > 0) \tag{3.78}$$

Notice that in the special case of $\beta = 1$, $f(t) = \alpha e^{-\alpha t}$, which is the density function of the exponential variable. The exponential variable is very seldom used in reliability studies, as it has the so-called forgetfulness property, which was already discussed earlier and is repeated here for convenience. If X is an exponential variable, then for all t, $\tau > 0$:

$$P(X > t + \tau \,|\, X > \tau) = P(X > t) \tag{3.79}$$

This relation shows that the probability of the working condition in any time length t is independent of how long the system was working before. Equation (3.79) can be shown as:

$$P(X > t + \tau \,|\, X > \tau) = \frac{P(X > t + \tau)}{P(X > \tau)} = \frac{R(t + \tau)}{R(\tau)} = \frac{e^{-\alpha(t+\tau)}}{e^{-\alpha\tau}} = e^{-\alpha t} = R(t) = P(X > t)$$

In most practical cases the breakdown probabilities increase in time as the system becomes older, so the exponential variable is inappropriate.

The *hazard-rate* is given as:

$$\rho(t) = \lim_{\Delta t \to 0} \frac{P(t \le X \le t + \Delta t \mid t \le X)}{\Delta t} \tag{3.80}$$

and shows how often failures will occur after time period t. It is clear that:

$$\rho(t) = \lim_{\Delta t \to 0} \frac{P(t \le X \le t + \Delta t)}{P(t \le X)\Delta t} = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \cdot \frac{1}{R(t)} = \frac{f(t)}{R(t)} \tag{3.81}$$

Therefore, the hazard rate can be computed as the ratio of the density function and the reliability function.

Example 3.22

Assume X follows the Weibull distribution. Then:

$$R(t) = 1 - F(t) = e^{-\alpha t^{\beta}}$$

Therefore, from Eq. (3.81) we have:

$$\rho(t) = \frac{\alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}}}{e^{-\alpha t^{\beta}}} = \alpha\beta t^{\beta-1} \tag{3.82}$$

which is an increasing polynomial of t showing that failure will occur more frequently for larger values of t. In the special case of an exponential distribution, $\beta = 1$ and $\rho(t) = \alpha$ is a constant.

An important problem in reliability engineering is to reconstruct $F(t)$ or the reliability function if the hazard rate is given, and we now offer a solution to this problem. Notice first that:

$$\rho(t) = \frac{f(t)}{1 - F(t)} = -\frac{\big(1 - F(t)\big)'}{1 - F(t)}$$

Integrate both sides in the interval [0,t] to obtain:

$$\int_0^t \rho(\tau)\, d\tau = \Big[-\ln\big(1 - F(\tau)\big)\Big]_{\tau=0}^{t} = -\ln\big(1 - F(t)\big) + \ln\big(1 - F(0)\big)$$

Because $F(0) = 0$, the second term equals zero, so:

$$\ln\big(1 - F(t)\big) = -\int_0^t \rho(\tau)\, d\tau$$

implying that

$$1 - F(t) = \exp\left(-\int_0^t \rho(\tau)\, d\tau\right)$$

and, finally,

$$F(t) = 1 - \exp\left(-\int_0^t \rho(\tau)\, d\tau \right) \tag{3.83}$$

Example 3.23

Assume first that the hazard rate is constant, with $\rho(t) = \alpha$. Then,

$$F(t) = 1 - \exp\left(-\int_0^t \alpha d\tau \right) = 1 - e^{-\alpha t}$$

showing that the distribution is exponential. Similarly, if $\rho(t)$ is a monomial, $\rho(t) = \alpha\beta t^{\beta-1}$, then:

$$F(t) = 1 - \exp\left(-\int_0^t \alpha\, \beta\, \tau^{\beta-1}\, d\tau \right) = 1 - \exp\left(-\left[\alpha\tau^{\beta} \right]_0^t \right) = 1 - e^{-\alpha t^{\beta}}.$$

demonstrating that the distribution is Weibull.

We know that $F(t)$ is increasing, $F(0) = 0$, and $\lim_{t\to\infty} F(t) = 1$. Similarly, $R(t)$ is decreasing, $R(0) = 1$, and $\lim_{t\to\infty} R(t) = 0$. The basic analytical properties of the hazard rate are now examined. If $\beta > 2$ and X is a Weibull variable, then $\rho(0) = 0$, $\lim_{t\to\infty} \rho(t) = \infty$, and $\rho(t)$ is a strictly increasing and strictly convex function; if $\beta = 2$, then $\rho(t)$ is linear; if $1 < \beta < 2$, then $\rho(t)$ is strictly increasing and strictly concave; if $\beta = 1$, then $\rho(t)$ is a constant; and if $\beta < 1$, then the hazard rate is decreasing in t, in which case defective systems tend to fail early. So, we have shown that the hazard rate decreases for a well-made system. Notice that:

$$p'(t) = \frac{f'(t)\left(1 - F(t)\right) + f(t)^2}{\left(1 - F(t)\right)^2}$$

which is positive if and only if the numerator is positive:

$$f'(t) > -\frac{f(t)^2}{1 - F(t)} = -f(t)\,\rho(t) \tag{3.84}$$

In this case, $\rho(t)$ is locally increasing; otherwise, $\rho(t)$ decreases.

The reliability of a component can be considered in the way shown here; however, the reliability of a system with several components depends on the configuration of the components in the entire system. Many systems are arranged in a series, some have a parallel configuration, and others are a combination of the two.

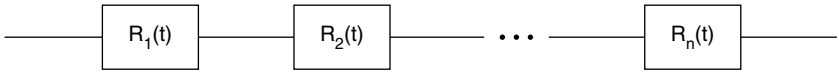


FIGURE 3.5 Series combination of components.

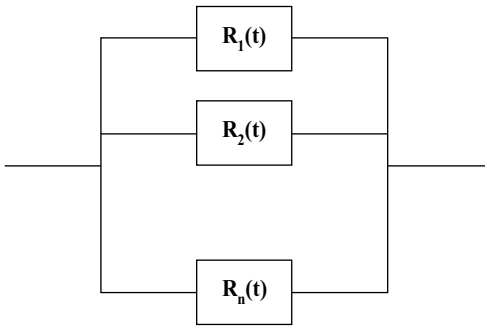


FIGURE 3.6 Parallel combination of components.

In a series system, the system fails if any of its components fails; in a parallel system, failure occurs when all components fail simultaneously.

Consider first the series combination shown in Figure 3.5. If $R_i(t)$ denotes the reliability function of component i ($i = 1, 2, \ldots, n$), then the reliability function of the system is given as:

$$R(t) = P(X > t) = P\Big(\big(X_1 > t\big) \cap \big(X_2 > t\big) \cap \cdots \cap \big(X_n > t\big)\Big)$$

where X is the time when the system fails, and $(X_1, \ldots, X_n)$ are the same for the components. If we assume that the failures of the different components occur independently of each other, then:

$$R(t) = P\big(X_1 > t\big)\, P\big(X_2 > t\big) \cdots P\big(X_n > t\big) = \prod_{i=1}^{n} R_i(t) \tag{3.85}$$

Notice that the inclusion of a new component in the system makes the reliability function smaller, as it is multiplied by the new factor, $R_{n+1}(t)$, which is less than one.

Consider next the parallel system of Figure 3.6. The system fails, if all components fail, so

$$P(X \le t) = P\Big(\big(X_1 \le t\big) \cap \big(X_2 \le t\big) \cap \cdots \cap \big(X_n \le t\big)\Big)$$

If the components fail independently of each other, then

$$R(t) = 1 - F(t) = 1 - P(X \le t) = 1 - \prod_{i=1}^{n} P\big(X_i \le t\big) = 1 - \prod_{i=1}^{n} F_i(t) = 1 - \prod_{i=1}^{n} \big(1 - R_i(t)\big) \tag{3.86}$$

where F and F_i are the cumulative distributions of time until the first failure of the system and component i, respectively. Notice that by including a new component, the reliability function increases, as the second term is multiplied by $1 - R_{n+1}(t)$, which is less than one.

In most cases, however, the connections between systems components are a mix of series and parallel. Then, Eqs. (3.85) and (3.86) are combined appropriately as is shown in the following example.

Example 3.24

Figure 3.7 shows a system assembled with five components, and components 2 and 3 are parallel. Using Eq. (3.86) we obtain:

$$R_2(t) = 1 - \prod_{i=1}^{3}\left(1 - R_2^i(t)\right)$$

and

$$R_3(t) = 1 - \left(1 - R_3^1(t)\right)\left(1 - R_3^2(t)\right)$$

Therefore, Eq. (3.85) implies that the reliability function of the system is as follows:

$$R(t) = R_1(t)\left[1 - \prod_{i=1}^{3}\left(1 - R_2^i(t)\right)\right]\cdot\left[1 - \prod_{i=1}^{2}\left(1 - R_3^i(t)\right)\right]\cdot R_4(t)\cdot R_5(t)$$

As a particular example, assume that:

$$R_1(t) = R_4(t) = R_5(t) = e^{-t}$$

and

$$R_2^i(t) = R_3^j(t) = e^{-2t} \quad (1 \le i \le 3, 1 \le j \le 2)$$

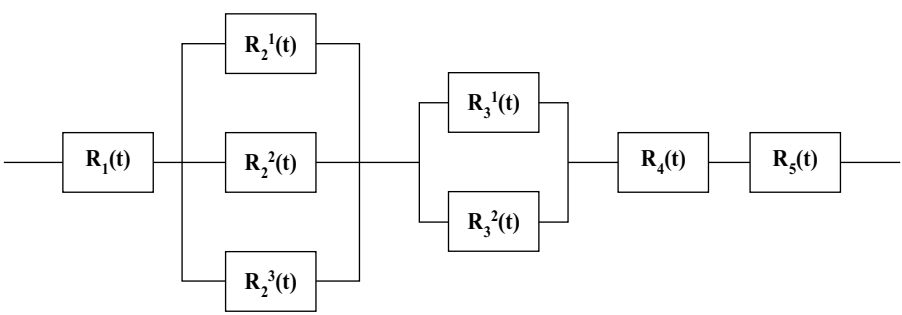


FIGURE 3.7 Combined connections in a system.

and the reliability of the system at $t = 0.1$ has to be determined. At $t = 0.1$,

$$R_1(0.1) = R_4(0.1) = R_5(0.1) = e^{-0.1} = 0.9048$$
$$R_2^i(0.1) = R_3^j(0.1) = e^{-0.2} = 0.8187$$

and we have:

$$R_2(0.1) = 1 - 0.1813^3 = 0.9940$$

and

$$R_3(0.1) = 1 - 0.1813^2 = 0.9671$$

Hence,

$$R(0.1) = 0.9048^3(0.9940)(0.9671) = 0.7121 = 71.21\%$$

We mention finally that Milton and Arnold (1995) provide much more details of the fundamentals of reliability theory.

3.5 STOCHASTIC PROCESSES

Suppose that a certain event occurs at random time points $0 \le t_1 < t_2 < \ldots$. These events constitute a *stochastic process*. For example, the times when earthquakes, floods, rainfalls, and so on occur represent stochastic processes. The process mathematically can be completely defined if we know the distribution functions of t_1, $t_2 - t_1$, $t_3 - t_2$, …. A very important characteristic of stochastic processes is the number of events, $N(t)$, that occur in the time interval $[0,t]$. The most frequently used stochastic process is the *Poisson process*, which is defined by the following properties:

1. $N(0) = 0$.
2. The number of events occurring in mutually exclusive time intervals are independent.
3. The distribution of the number of events occurring in any given time interval depends on only the length of the interval and not on its location.
4. $$\lim_{\Delta t \to 0} \frac{P(N(\Delta t) = 1)}{\Delta t} = \lambda \tag{3.87}$$
   where $\lambda > 0$ is a constant.
5. $$\lim_{\Delta t \to 0} \frac{P(N(\Delta t) \ge 2)}{\Delta t} = 0 \tag{3.88}$$

Condition (1) shows that the process starts at $t = 0$. If $0 < t_1 < t_2 < t_3 < t_4$, then condition (2) states that the number of events in interval $[t_1,t_2]$, which is $N(t_2) - N(t_1)$, is independent of the number of events in interval $[t_3,t_4]$, which is $N(t_4) - N(t_3)$. This condition is called the *independent increment assumption*. Condition (3) shows that the distribution of $N(t + s) - N(t)$ depends on only s and is independent of t. This condition is known as the *stationary increment assumption*. Conditions (4) and (5) state that in a small interval of length Δt, the probability that one event occurs is approximately $\lambda\Delta t$, but the probability that at least two events occur is approximately zero.

Under conditions (1) to (5), for $k = 0, 1, 2, \ldots$,

$$P\big(N(t) = k\big) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{3.89}$$

That is, $N(t)$ has a Poisson distribution with parameter λt. The random variables

$$X_1 = t_1, X_2 = t_2 - t_1, X_3 = t_3 - t_2, \cdots$$

represent the sequence of inter-arrival times. Let their distribution functions be denoted by $F_1, F_2, F_3, \ldots$. Then, clearly,

$$F_1(t) = P(t_1 < t) = 1 - P(t_1 \geq t) = 1 - P(N(t) = 0) = 1 - e^{-\lambda t} \tag{3.90}$$

from Eq. (3.89); therefore, X_1 is exponentially distributed with expectation $1/\lambda$. For all $k \geq 2$ and $s, t \geq 0$,

$$\begin{aligned} P(X_k > t \,|\, X_{k-1} = s) &= P(0 \text{ event occurs in} (s, s+t) \,|\, X_{k-1} = s) \\ &= P(0 \text{ event occurs in} (s, s+t)) \\ &= P(0 \text{ event in} (0,t)) = P(N(t) = 0) = e^{-\lambda t} \end{aligned} \tag{3.91}$$

so the distribution function of X_k is the same as that of X_1 and is given in Eq. (3.90). Hence, all variables $X_1, X_2, X_3, \ldots$ are exponential with the same parameter λ, and they are independent of each other. More details on the Poisson process are offered, for example, in Ross (1987).

3.6 MARKOV CHAINS

Consider a dynamic system with discrete time scales, $t = 0, 1, 2, \ldots$. Assume that at each time period the state of the system can be any one from the finite set $\{S_1, S_2, S_3, \ldots, S_n\}$, and at each time period the state changes randomly. Let $X(t)$ denote the state at time period t and assume that the *transition probabilities*:

$$P_{ij} = P\Big\{X(t+1) = S_j \,\Big|\, X(t) = S_i\Big\} \tag{3.92}$$

are independent of t. Notice that P_{ij} is the probability that the state at time period $t+1$ is S_i given that the state in the previous time period t is S_i. It is convenient to form the *transition matrix*:

$$\underline{P} = \begin{pmatrix} P_{11} & P_{12} \cdots P_{1n} \\ P_{21} & P_{22} \cdots P_{2n} \\ \vdots & \vdots \quad \vdots \\ P_{n1} & P_{n2} \cdots P_{nn} \end{pmatrix} \tag{3.93}$$

which has the following fundamental properties:

1. $0 \le P_{ij} \le 1$, for all i and j.
2. $\sum_{j=1}^{n} P_{ij} = 1$ for all i.

Any matrix satisfying these conditions is called a *stochastic matrix*.

At any time period t, let $p_i(t)$ denote the probability that the system is in state S_i:

$$p_i(t) = P\big(X(t) = S_i\big) \tag{3.94}$$

Clearly,

3. $0 \le p_i(t) \le 1$ for all i and t;
4. $\sum_{i=1}^{n} p_i(t) = 1$ for all t.

Next, define vector $\underline{p}(t)^T = (p_1(t), p_2(t), ..., p_n(t))$ as a row vector. Any n-element vector satisfying these two conditions is called a *probability vector*. Notice that $p_i(t)$ gives the probability that the system is in state S_i at time period t. By using the transition matrix $\underline{P}$, we are able to establish a direct relation between the consecutive state-probability vectors $\underline{p}(t)$ and $\underline{p}(t+1)$. By using the well-known idea of conditioning, it is easy to see that:

$$\begin{aligned} p_j(t+1) = P\big(X(t+1) = S_j\big) &= \sum_{i=1}^{n} P\big(X(t+1) = S_j \big| X(t) = S_i\big) P(X(t) = S_i) \\ &= \sum_{i=1}^{n} P_{ij} p_i(t) \end{aligned} \tag{3.95}$$

showing that

$$\underline{p}(t+1)^T = \underline{p}(t)^T \underline{P} \tag{3.96}$$

or by taking the transpose of both sides:

$$\underline{p}(t+1) = \underline{P}^T \underline{p}(t) \tag{3.97}$$

Notice that this is a homogeneous, discrete, linear system with state-vector $\underline{p}(t)$ at time period t, so all methods known from systems theory can be applied here.

Let $\underline{p}\,(0)^T$ be the probability vector of the initial time period. Then, by the repeated application of Eq. (3.96), we obtain:

$$\underline{p}(1)^T = \underline{p}(0)^T \underline{P},$$
$$\underline{p}(2)^T = \underline{p}(1)^T P = \underline{p}(0)^T \underline{P}^2$$

and so on, and in general:

$$\underline{p}(t)^T = \underline{p}(0)^T \underline{P}^t \tag{3.98}$$

for all $t \geq 1$.

Example 3.25

Assume that the weather in a city can be characterized on any day as sunny (S_1), cloudy (S_2), or rainy (S_3). If it is sunny one day, then the probability that it will remain sunny or become cloudy is 50% for either. If it is cloudy one day, then the probability that the next day will be sunny is 40%; cloudy, 30%; and rainy, 30%. If it is rainy one day, then the probability that it will be rainy again the next day is 50%; cloudy, 40%; and sunny, 10%. Based on these probability values we can construct the transition matrix:

$$\underline{P} = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.4 & 0.3 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$$

Assume that one day the weather is sunny; that is,

$$\underline{p}(0)^T = (1, 0, 0)$$

Then, from Eq. (3.96), with $t = 0$ we obtain:

$$\underline{p}(1)^T = (1,0,0)\begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.4 & 0.3 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{pmatrix} = (0.5, 0.5, 0)$$

$$\underline{p}(2)^T = (0.5,0.5,0)\begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.4 & 0.3 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{pmatrix} = (0.45, 0.40, 0.15)$$

and so on. For example, the components of $\underline{p}(2)^T$ show that two days later the probability that the weather will be sunny is 45%; cloudy, 40%; and rainy, 15%.

A Markov chain is *regular*, if $\underline{P}^t > \underline{0}$ for some possible integer t. In the previous example,

$$\underline{p}^2 = \begin{pmatrix} 0.45 & 0.4 & 0.15 \\ 0.35 & 0.41 & 0.24 \\ 0.26 & 0.37 & 0.37 \end{pmatrix} > \underline{0}$$

Thus, our weather example is described by a regular Markov chain. Regular Markov chains satisfy the following condition:

5. There is a unique probability vector $\underline{p}^T$ such that:

$$\underline{p}^T = \underline{p}^T \underline{P}. \tag{3.99}$$

   which is called the *long-term probability vector* of the system.
6. For arbitrary initial state probability vector $\underline{p}(0)^T$,

$$\lim_{t\to\infty} \underline{p}(0)^T \underline{P}^t = \underline{p}^T \tag{3.100}$$

   That is, the left-hand side limit always exists and equals $\underline{p}^T$.
7. The limit:

$$\lim_{t\to\infty} \underline{P}^t = \overline{\underline{P}} \tag{3.101}$$

   is an $n \times n$ matrix for which each of the rows is equal to $\underline{p}^T$.

Equation (3.99) and the additional equation:

$$\sum_{i=1}^{n} p_i = 1$$

provide a system of linear equations to determine the long-term probability vector. Notice that Eq. (3.99) consists of n linear equations, which are linearly dependent, as by adding the components of both sides we have:

$$\underline{p}^T \underline{1} = 1$$

and

$$\underline{p}^T \underline{P1} = \underline{p}^T (\underline{P1}) = \underline{p}^T \underline{1} = 1$$

The additional equation serves as the nth required independent equation in order to guarantee a unique solution.

Example 3.26

In the previous example, Eq. (3.99) and the additional equation give the linear system:

$$\begin{aligned} p_1 &= 0.5p_1 + 0.4p_2 + 0.1p_3 \\ p_2 &= 0.5p_1 + 0.3p_2 + 0.4p_3 \\ p_3 &= \quad\quad 0.3p_2 + 0.5p_3 \\ p_1 &+ p_2 + p_3 = 1. \end{aligned}$$

It is easy to see that the unique solution is the following:

$$p_1 = \frac{23}{63} \approx 0.365, \quad p_2 = \frac{25}{63} \approx 0.397, \quad p_3 = \frac{15}{63} \approx 0.238$$

The fundamentals of Markov chains and special cases are discussed, for example, in Luenberger (1979).

3.7 OPTIMIZATION UNDER UNCERTAINTY

In hydrology, the consequences of our decisions as well as our options depend on random elements not known at the time when the decisions are made. In this section, two popular methodologies will be discussed that work well to solve optimization problems with random and uncertain factors.

3.7.1 Chance-Constrained Programming

Consider an optimization problem for which both the objective function and the constraints depend on a random variable:

$$\begin{aligned} &\text{Maximize} \quad f\left(\underline{x}, \underline{\xi}\right) \\ &\text{Subject to} \quad \underline{g}(\underline{x}, \underline{\xi}) \geq \underline{0} \end{aligned} \tag{3.102}$$

where $\underline{x} \in R^n$ is the decision vector, $\underline{\xi}$ is a random variable, f is a real-valued function, and $\underline{g} \in R^m$. It is usual to replace the stochastic objective function by its expectation with respect to $\underline{\xi}$; however, this expectation-optimizing concept itself does not require the feasibility of the optimal solution. For any decision $\underline{x}$, there is always a chance that it is not feasible for certain values of $\underline{\xi}$. Therefore, the constraints are substituted by the requirements that the probability of infeasibility has to be below a certain tolerance level. This condition can be posed on each individual constraint or on the entire constraint set. In the case of infeasibility, losses might occur, and their expectations are usually added to the expected value of the original objective function. Therefore, we now consider the following problem:

$$\begin{aligned} &\text{Maximize} \quad E\left[f\left(\underline{x}, \underline{\xi}\right)\right] - E\left[L\left(\underline{g}\left(\underline{x}, \underline{\xi}\right)\right)\right] \\ &\text{Subject to} \quad P\left(\underline{g}\left(\underline{x}, \underline{\xi}\right) \geq \underline{0}\right) \geq 1 - \varepsilon \end{aligned} \tag{3.103}$$

where L is a loss function with zero value if the constraints are all satisfied. This problem is now a deterministic nonlinear optimization problem that can be solved by standard methodology. Let g_i denote the ith component of $\underline{g}$. Then, alternatively, the constraint of Eq. (3.103) might be replaced by the set of constraints:

$$P\left(g_i\left(\underline{x}, \underline{\xi}\right) \geq 0\right) \geq 1 - \varepsilon_i \quad (\text{for all } i) \tag{3.104}$$

We now introduce the notation:

$$F(\underline{x}) = \left\{\underline{\xi} \middle| \underline{g}\left(\underline{x}, \underline{\xi}\right) \geq \underline{0}\right\} \tag{3.105}$$

and let p denote the density function of $\underline{\xi}$. Then, the constraint of Eq. (3.103) can be rewritten as:

$$\int_{F(\underline{x})} p\left(\underline{\xi}\right) d\underline{\xi} \geq 1 - \varepsilon \tag{3.106}$$

and, similarly, the objective function is also an integral:

$$\int \left[f\left(\underline{x}, \underline{\xi}\right) - L\left(\underline{g}\left(\underline{x}, \underline{\xi}\right)\right) \right] p\left(\underline{\xi}\right) d\underline{\xi} \tag{3.107}$$

In practical applications, numerical integration is used in computing the objective function values (see, for example, Szidarovszky and Yakowitz, 1978). A good survey of stochastic optimization methodology is offered in Wets (1983).

3.7.2 Stochastic Dynamic Programming

In the case of DP formulation, we assumed that the state transition relation:

$$x_{t+1} = f_t(x_t, u_t) \tag{3.108}$$

uniquely determines the next state given the current state and decision. In the stochastic version (SDP), this deterministic function is replaced by a probabilistic transition function $p_t(x,x_t,u_t)$, which gives the probability density function of x_{t+1} for fixed t, x_t, and u_t. Because of the presence of stochastic elements, a decision strategy is a sequence of function $u_t = D_t(x)$, which selects the decision depending on the particular state $x_t = x$ occurring at time period t. We also replace the deterministic objective function by its expectation:

$$E\left[\sum_{t=0}^{T} g_t\left(x_t, x_{t+1}, D_t(x_t)\right) \right] \tag{3.109}$$

where we assume that g_t also depends on the next state. In the deterministic case, x_{t+1} is uniquely determined by x_t and u_t, so x_{t+1} would be a superfluous variable. The stochastic version of the DP recursive equation is the following:

$$V_t(x) = \max_u \left[\int \left(g_t(x, x', u) + V_{t+1}(x') \right) \right] p_t(x', x, u) dx' \tag{3.110}$$

with the choice of $V_{T+1}(x) = 0$ for all x. Let $u = S_t(x)$ denote an optimal choice. Then, $D_t(x) = S_t(x)$ is optimal for objective (3.109). The corresponding discrete stochastic dynamic programming (DSDP) algorithm is based on the recursion:

$$\hat{V}_t\left(x^{(i)}\right) = \min_j \left\{ \sum_k g_t\left(x^{(i)}, x^{(k)}, u^{(j)}\right) + \hat{V}_{t+1}\left(x^{(k)}\right) P_t\left(x^{(k)}, x^{(i)}, u^{(j)}\right) \right\} \tag{3.111}$$

where P_t is the probability mass function obtained from the density function p_t and the discretization scheme being used. The curse of dimensionality can be made less

burdensome by using some version of DDDP or IDP. The survey article of Yakowitz (1982) has more details of the methodology discussed here and applications in to water resources management.

3.8 RISK, UNCERTAINTY, AND BAYESIAN DECISION THEORY

Decision problems under uncertainty are usually formulated as:

$$\begin{aligned} &\text{Maximize} \quad f(x,\xi) \\ &\text{Subject to} \quad x \in X(\xi) \end{aligned} \tag{3.112}$$

where ξ is the notation for the uncertain phenomenon. Notice that both the objective function and the feasible set depend on ξ. First, we show that the dependence of $X(\xi)$ on ξ can be formally eliminated. For this purpose, we define:

$$X = \underset{\xi \in \sum}{U} X(\xi) \tag{3.113}$$

where $\sum$ is the set of all possible values of ξ, and we let the modified objective function be:

$$g(x,\xi) = \begin{cases} f(x,\xi) & \text{if} \quad x \in X(\xi) \\ -\infty & \text{otherwise.} \end{cases} \tag{3.114}$$

Thus, the problem can be reformulated as:

$$\begin{aligned} &\text{Maximize} \quad g(x,\xi) \\ &\text{Subject to} \quad x \in X. \end{aligned} \tag{3.115}$$

In the *classical approach*, the maxmin rule is most frequently applied. For each decision alternative x we compute the worst possible outcome:

$$g_w(x) = \inf_{\xi} g(x,\xi) \tag{3.116}$$

Then, the alternative with the highest worst outcome, x^*, is selected as the decision:

$$g_w(x^*) = \sup_{x} g_w(x) = \sup_{x} \inf_{\xi} g(x,\xi) \tag{3.117}$$

This solution concept is based on worst-case analysis in which decision makers try to protect themselves as much as possible against the worst possibility that might

occur. In many cases, worst possibilities occur only very seldom, so protection against such events seems to be overprotection in most cases. Another approach to choose a single decision from X is to assign a weight function $w(\xi)$ over all possible values of ξ and choose the decision that maximizes:

$$\int_{\Sigma} g(x,\xi)w(\xi)d\xi \tag{3.118}$$

and if Σ contains only finite or countable many values, then the integral is replaced by the summation:

$$\sum_{\xi} g(x,\xi)w(\xi) \tag{3.119}$$

It seems reasonable to assign a particular weight function $w(\xi)$ that reflects the relative likelihood of the different values in Σ being the true value of ξ. With this reasoning, we have revealed the basic concept of statistical decision making.

3.8.1 Risk and Value of Information

In statistical decision making, we always assume that the uncertainty in ξ is based on its randomness. Let h denote the density function of ξ (or the joint density function if ξ has more than one component). According to Eqs. (3.118) or (3.119), the expected value of the objective function is maximized:

$$\begin{aligned} &\text{Maximize} \quad E\left[g(x,\xi)\right] \\ &\text{Subject to} \quad x \in X. \end{aligned} \tag{3.120}$$

Let x^* be the optimal solution. For each particular value of ξ let $x(\xi)$ be the maximizer of $g(x,\xi)$. Because the true value of ξ is unknown, we cannot select the corresponding optimal decision $x(\xi)$; instead, x^* is the choice. This lack of information results in the loss (or we face with risk) of:

$$L(\xi) = g\left(x(\xi),\xi\right) - g\left(x^*,\xi\right) \tag{3.121}$$

for each particular value of ξ. Because the true value of ξ is not known, we can compute only its expectation:

$$E\left[L(\xi)\right] = E\left[g\left(x(\xi),\xi\right)\right] - E\left[g\left(x^*,\xi\right)\right] \tag{3.122}$$

which gives the average loss occurring by not knowing the value of ξ. In other words, this is the value of the knowledge of the true value of ξ, or it is the value of the perfect information on ξ.

Most decisions in water resources management are made repeatedly over time. As time progresses, new information about the uncertain factors is gained and their distributions can be updated so more accurate decisions can be made. The methodology used in such situations is known as *Bayesian decision making*, the elements of which are the subject of the next subsection.

3.8.2 Bayesian Decision Making

Let again ξ denote the uncertain model parameter. Before additional information is available its density function is believed to be $h(\xi)$. This density function is called the *prior distribution*, because it is the distribution of ξ before additional observations are made. Let θ be a random variable or vector that is related to the parameter ξ, and assume that a value of θ is being observed. Then, the *posterior* density of ξ can be obtained by using Bayes' theorem:

$$h(\xi|\theta) = \frac{l(\theta|\xi)h(\xi)}{\int_{\Sigma} l(\theta|\xi')h(\xi')d\xi'} \tag{3.123}$$

where $l(\theta \mid \xi)$ is the conditional density function of θ given the value of ξ. If ξ has a discrete distribution, then the denominator of the right-hand side is replaced by a sum. So, after observing θ, the density function of ξ is updated and a more accurate decision can be made. The benefit of the additional information θ is given by the difference:

$$g(x^{**},\xi) - g(x^{*},\xi) \tag{3.124}$$

where x^{**} is the optimal decision based on the updated density function $h(\xi \mid \theta)$. Because ξ is unknown and an updated density is available, the expected benefit is computed as:

$$\int_{\Theta}\int_{\Sigma} \left[g(x^{**},\xi) - g(x^{*},\xi)\right]h(\xi|\theta)d\xi\bar{l}(\theta)d\theta \tag{3.125}$$

where $\bar{l}(\theta)$ is the denominator of the right-hand side of Eq. (3.123), and Θ is the set of the values of θ. Based on Eq. (3.125), we can decide if it is worthwhile to wait for an additional observation by comparing the expected benefit to the cost of collecting the additional observation and any other losses occurring by delaying the decision.

3.8.3 Conjugate Distribution Families

Assume that an observation is made on a random variable or vector θ as was assumed in the previous section. Usually, it is a random sample from one of the standard distributions such as Bernoulli, binomial, or Poisson. All of these distributions depend on certain parameters. For the Bernoulli variable, the parameter is p; for binomial, n and p; for Poisson, λ; and so on. Let π denote the parameter or parameter vector of the distribution in general. A convenient property of a family of prior distributions is that it is *closed under sampling*, which means that, if the prior distribution belongs to the family, then for any sample size and observation values the posterior distribution also belongs to the same family. A distribution family with this property is a *conjugate family of prior distributions*.

In selecting a conjugate family in a particular application we have to consider the following. The family has to be small enough so it can be described by only a few parameters, so computationally it is tractable. The family also has to be rich enough to allow the decision makers to find within the prior family a distribution that fits adequately the prior distribution.

Assume first that ξ has a Bernoulli distribution with possible values 1 and 0 and occurring probabilities $P(\xi = 1) = \pi$ and $P(\xi = 0) = 1 - \pi$. The most commonly used conjugate family for π is the family of *beta distributions* with density function:

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1} \quad (0 \le \pi \le 1) \tag{3.126}$$

where α and β are constants. Let $\theta_1, \theta_2, \ldots, \theta_n$ be a random sample of the Bernoulli distribution and assume that its parameter π follows a beta distribution with constants α and β. Then, the posterior distribution of π given the observed values $\theta_1, \theta_2, \ldots, \theta_n$ is also a beta distribution with constants:

$$\alpha + \sum_{i=1}^{n} \theta_i, \beta + n - \sum_{i=1}^{n} \theta_i$$

Assume next that ξ is normally distributed with mean m and precision p (which is $1/\sigma^2$), so its density function is:

$$\sqrt{\frac{p}{2\pi}} e^{-\frac{p(\xi-m)^2}{2}} \qquad (-\infty < \xi < \infty) \tag{3.127}$$

Assume that precision p is known and the prior distribution of m is normal with mean μ and precision π. If $\theta_1, \theta_2, \ldots, \theta_n$ is a random sample of ξ, then the posterior distribution of m is also normal with mean:

$$\mu' = \frac{\pi\mu + p\sum_{i=1}^{n}\theta_i}{\pi + np}$$

and precision $\pi' = \pi + np$.

The conjugate family becomes more complicated when both m and p are unknown. In this case, we have the following result. Assume that ξ is normally distributed with mean m and precision p, both of which are unknown, and their joint distribution is as follows: The conditional distribution of m given the value of p is normal with mean μ and precision τp, where τ is a positive constant and the marginal distribution of p is gamma with parameters α and β; that is, the marginal density function of p is given as:

$$\frac{\beta^{\alpha}}{\Gamma(\alpha)} p^{\alpha-1} e^{-\beta p} \qquad (p > 0) \tag{3.128}$$

Let $\theta_1, \theta_2, \ldots, \theta_n$ be a random sample of ξ. Then, the posterior joint distributions of m and p are as follows. The conditional distribution of m given the value of p is again a normal distribution with mean:

$$\frac{\tau\mu + \sum_{i=1}^{n}\theta_i}{\tau + n}$$

and precision $(\tau + n)p$, and the marginal distribution of p is again gamma with parameters

$$\alpha' = \alpha + \frac{n}{2} \quad \text{and} \quad \beta' = \beta + \frac{1}{2}\sum_{i=1}^{n}\left(\theta_i - \bar{\theta}\right)^2 + \frac{\tau n\left(\bar{\theta} - \mu\right)^2}{2(\tau + n)}$$

where

$$\bar{\theta} = \frac{1}{n}\sum_{i=1}^{n}\theta_i$$

The most commonly used conjugate families other than the ones presented above are given, for example, in De Groot (1970).

3.9 FUZZY SETS AND PARAMETER IMPRECISION

Parameter uncertainty in water resources and hydrologic modeling is handled by randomization or by fuzzification. The previous section discussed randomization and the corresponding probabilistic methods; in this section, the fundamentals of fuzzy sets and fuzzy decisions are introduced.

Let X be the set of certain objects. A *fuzzy set* A in X is a set of ordered pairs:

$$A = \left\{\left(x, \mu_A(x)\right)\right\}, \quad x \in X \tag{3.129}$$

where $\mu_A : X \mapsto [0,1]$ is called the membership function and $\mu_A(x)$ is the grade of membership of x in A. In classical set theory, x either belongs to A or does not belong to A, so $\mu_A(x)$ equals 1 or 0, respectively.

The basic relations and operations of fuzzy sets can be defined as follows:

1. A fuzzy set A is *empty*, if $\mu_A(x) = 0$ for all $x \in X$.
2. A fuzzy set A is called *normal*, if:

$$\sup_x \mu_A(x) = 1 \tag{3.130}$$

   Notice that any nonempty fuzzy set can be normalized by introducing the normalized membership function:

$$\overline{\mu}_A(x) = \frac{\mu_A(x)}{\sup_x \mu_A(x)} \tag{3.131}$$

3. The *support* of a fuzzy set A is defined as:

$$S(A) = \left\{x \middle| x \in X, \mu_A(x) > 0\right\} \tag{3.132}$$

4. We say that fuzzy sets A and B are equal, if:

$$\mu_A(x) = \mu_B(x) \ \text{(for all } x \in X) \tag{3.133}$$

5. A fuzzy set A is a *subset* of B, if for all $x \in X$,

$$\mu_A(x) \le \mu_B(x) \tag{3.134}$$

   It is clear that $A=B$ if A is a subset of B and B is also a subset of A. The relation that A is a subset of B is denoted as $A \subseteq B$.

6. A' is the *complement* of A, if for all $x \in X$:

$$\mu_{A'}(x) = 1 - \mu_A(x) \tag{3.135}$$

Notice that this is the straightforward generalization of the concept of complement of classical set theory, as $x \in A$ if and only if $\mu_A(x) = 1$ and $x \notin A'$, so $\mu_{A'}(x) = 0$. Similarly, $x \notin A$ if and only if $\mu_{A'}(x) = 1$ and $\mu_A(x) = 0$.

7. The *intersection* of fuzzy sets A and B is defined as:

$$\mu_{A\cap B}(x) = \min\{\mu_A(x); \mu_B(x)\} \quad (\text{for all } x \in X) \tag{3.136}$$

Notice that conditions (5) and (7) imply that $A \subseteq B$ if and only if $A \cap B = A$.

8. The *union* of fuzzy sets A and B is given as:

$$\mu_{A\cup B}(x) = \max\{\mu_A(x); \mu_B(x)\} \quad (\text{for all } x \in X) \tag{3.137}$$

From conditions (5) and (8), we see that $A \subseteq B$ if and only if $A \cup B = B$.

9. The *algebraic product* of fuzzy sets A and B is denoted by AB and is defined by the relation:

$$\mu_{AB}(x) = \mu_A(x)\mu_B(x) \ (\text{for all } x \in X) \tag{3.138}$$

10. The *algebraic sum* of fuzzy sets A and B is denoted by $A + B$ and has the membership function:

$$\mu_{A+B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x) \quad (\text{for all } x \in X) \tag{3.139}$$

11. A fuzzy set A is *convex*, if for all $x, y \in X$ and $\lambda \in [0,1]$:

$$\mu_A(\lambda x + (1-\lambda)y) \geq \min\{\mu_A(x); \mu_A(y)\} \tag{3.140}$$

If A and B are convex, then it is easy to see that $A \cap B$ is also convex.

12. A fuzzy set A is *concave*, if A' is convex. It is easy to see that if A and B are concave, then $A \cup B$ is also concave.

13. Let X and Y be two sets and $f: X \mapsto Y$ be a mapping from X to Y. Assume that A is a fuzzy set in X. The fuzzy set B induced by mapping f is defined in Y with membership function:

$$\mu_B(y) = \sup_{x \in f^{-1}(y)} \mu_A(x) \tag{3.141}$$

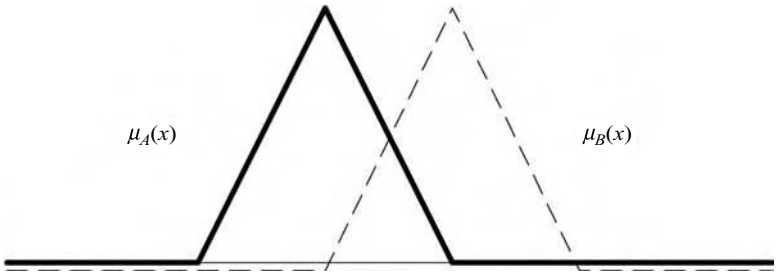


FIGURE 3.8 Membership functions of A and B.

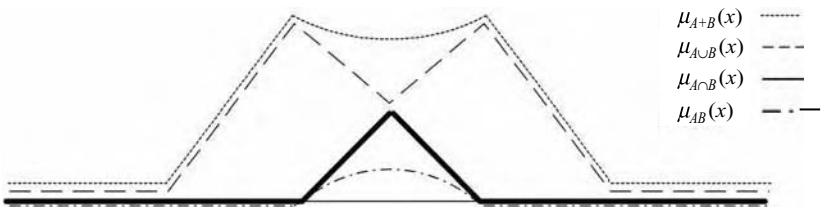


FIGURE 3.9 Operations of fuzzy sets $A \cap B$, $A \cup B$, AB, $A + B$.

where

$$f^{-1}(y) = \{x | x \in X, f(x) = y\}$$

Example 3.27

Let X = [–∞,∞] and fuzzy sets A and B be defined by the membership functions shown in Figure 3.8. The membership functions of $A \cap B$, $A \cup B$, AB, and $A + B$ are presented in Figure 3.9.

Example 3.28

Let $X = [-1,1]$ and assume that

$$\mu_A(x) = \begin{cases} 1+x & \text{if} \quad -1 \le x \le 0 \\ 1-x & \text{if} \quad 0 < x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $f(x) = x^2$. We will now determine the membership function of the fuzzy set induced by function f. Notice first that $f([-1,1]) = [0,1]$, so $\mu_B(y) = 0$ if $y \notin [0,1]$. If $y \in [0,1]$, then

$$f^{-1}(y) = \{\sqrt{y}, -\sqrt{y}\}$$

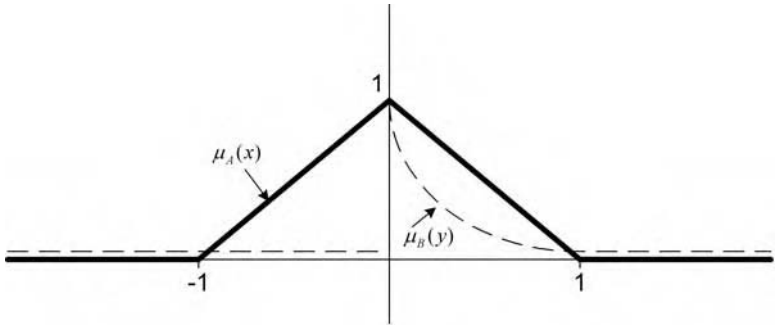


FIGURE 3.10 Illustration of $\mu_A(x)$ and $\mu_B(y)$.

and because

$$\mu_A(\sqrt{y}) = 1 - \sqrt{y} = \mu_A(-\sqrt{y}),$$
$$\mu_B(y) = \sup_{x \in f^{-1}(y)} \mu_A(x) = 1 - \sqrt{y}.$$

Both $\mu_A(x)$ and $\mu_B(y)$ are illustrated in Figure 3.10.

Uncertain model parameters can be considered as fuzzy sets, and the result of any series of operations is a fuzzy set. Uncertainty in constraints and in the objective functions also can be expressed by fuzzy sets.

Let X be the set of decision alternatives. A fuzzy constraint is defined by a fuzzy set C in X, and the membership function value $\mu_C(x)$ shows to what degree an alternative satisfies the constraint C. Similarly, any objective function can be identified by a fuzzy set G, and $\mu_G(x)$ shows to what degree an alternative represents an attainment of the goal represented by this objective function. Any multiple-criteria decision-making problem can be defined in a fuzzy environment as the set of fuzzy goals (G_1, …, G_M) and fuzzy constraints (C_1, …, C_N). The solution of the problem is the decision alternative that satisfies the constraints C_j to as high a level as possible and gives the best possible attainment of the goals. In other words, the intersection of the fuzzy sets (G_1, …, G_M) and (C_1, …, C_N) must have the highest membership value. Therefore, we consider:

$$\mu_D(x) = \min\left\{\mu_{G_1}(x), \cdots, \mu_{G_M}(x); \mu_{C_1}(x), \cdots, \mu_{C_N}(x)\right\} \tag{3.142}$$

and find alternative x that maximizes this membership function.

Example 3.29

Assume we have six alternatives, two constraints, and two objectives. The membership values for all alternatives with respect to the constraints and objectives are given

TABLE 3.2
Membership Values in Fuzzy Decision

	Alternatives					
	1	**2**	**3**	**4**	**5**	**6**
μ_{G_1}	0.9	0.85	0.95	0.8	0.65	0.9
μ_{G_2}	0.95	0.70	0.92	0.9	0.75	0.95
μ_{C_1}	0.75	0.8	0.9	0.95	0.9	0.75
μ_{C_2}	0.8	0.9	0.91	0.85	0.8	0.8
μ_D	0.75	0.7	0.9	0.8	0.65	0.75

in Table 3.2. The last row gives the values of μ_D by selecting the smallest numbers of all columns. The largest value of μ_D appears in the third column, so the third alternative is the best.

Interested readers are referred to Bellman and Zadeh (1970) or to the section on fuzzy decisions in Szidarovszky et al. (1986).

3.10 PROBLEMS

3.1 Assume that the number of occurrences in time period [0,1] has Poisson distribution with expectation $\lambda = 3$. Find the probability that the first occurrence will happen before $t = 1/2$.

3.2 A density function of a random variable X is given as:

$$f(x) = \begin{cases} a(x+1) & \text{if} \quad 0 < x < 2 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of a.
(b) Find $F(X)$.
(c) Compute $E(X)$ and $Var(X)$.

3.3 Find the mode and the median of the random variable of the previous problem.

3.4 A discrete random variable takes on the values 0, 1, 2, 3 with probability $P(X = k) = a/k^2$ ($k \geq 1$) and $P(X = 0) = 2a$, where $a > 0$ is a parameter.
(a) Find the value of a.
(b) Find $E(X)$ and $Var(X)$.
(c) Find the mode and median.

3.5 Find the moment generating function of the random variable of problem 3.2.

3.6 A joint probability table is given below:

	Y	
X	**0**	**1**
0	0.1	0.3
1	0.1	0.2
2	0.2	*A*

(a) Find the value of A.
(b) Are X and Y independent?
(c) Compute $E(X)$, $E(Y)$, $Var(X)$, and $Var(Y)$.
(d) Compute $Cov(X,Y)$ and $Corr(X,Y)$.

3.7 A joint density function is given as:

$$f(x,y)=\begin{cases}A(x+y) & \text{if} \quad 0<x<y<1\\ 0 & \text{otherwise.}\end{cases}$$

Solve the previous problem for this continuous case.

3.8 Consider a neural network with two input nodes, two output nodes, and three hidden nodes. Generate data by selecting the input values x_1 and x_2 from the unit interval [0,1] and using the output functions $y_1 = x_1^2 + x_2^2$ and $y_2 = x_1^2 - x_2^2$. Use back propagation to train the network.

3.9 Construct an observer for system:

$$\underline{\dot{x}} = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}\underline{x} + \begin{pmatrix}1\\ 2\end{pmatrix}u$$

$$y = (1,1)\,\underline{x}$$

Select matrix $\underline{K}$ such that the eigenvalues of the feedback system become –1 and –2.

3.10 Consider the distribution function of a random variable X:

$$F(x)=\begin{cases}0 & \text{if} \quad x\le 0\\ x^2 & \text{if} \quad 0<x\le 1\\ 1 & \text{if} \quad x\le 1\end{cases}$$

Derive an algorithm to generate random values of X from a uniform variable.

3.11 Solve the previous problem if:

$$F(x)=\begin{cases}1-\dfrac{1}{x} & \text{if} \quad x>1\\ 0 & \text{if} \quad x\le 1\end{cases}$$

3.12 Find the reliability functions and hazard rates for the random variables of problems 3.10 and 3.11.

3.13 A parallel system contains 10 elements with common hazard rate $\rho(t) = 0.1$. Find the hazard rate of the system.

3.14 Solve the previous problem if the system is a series combination of 10 same elements.

3.15 Consider a Poisson process with $\lambda = 2$. Find the probability that there are exactly three events before $t = 2$.

3.16 Find the long-term probability vector if the probability for the Markov chain with transition matrix is:

$$\underline{P}=\begin{pmatrix}0.2 & 0.6 & 0.2\\ 0.4 & 0.4 & 0.2\\ 0 & 0.6 & 0.4\end{pmatrix}$$

3.17 Assume that the initial probabilities are:

$$P_1(0)=0.1,\ P_2(0)=0.6,\ \text{ and }\ P_3(0)=0.3$$

Find the probabilities $P_i(3)$, $i = 1, 2, 3$.

3.18 Consider the following linear programming problem:

$$\begin{aligned}&\text{Maximize} && x_1+x_2\\ &\text{Subject to} && x_1,x_2\ge 0\\ & && 2x_1+x_2\le \xi\\ & && x_1+2x_2\le 3\end{aligned}$$

where ξ is normally distributed with $\mu = 3$ and $\sigma^2 = 1$. Give the chance constraint formulation of the problem, and find the optimal solution.

3.19 Consider the following optimization problem:

$$\begin{aligned}&\text{Maximize} && x_1+\xi x_2\\ &\text{Subject to} && x_1,x_2\ge 0\\ & && x_1+2x_2\le 4\\ & && 2x_1+x_2\le 4\end{aligned}$$

where ξ is a uniform random variable in [0,1]. Formulate Eq. (3.120) and find the optimal solution.

3.20 Repeat Example 3.29 for the following fuzzy decision problem:

	Alternatives					
	1	**2**	**3**	**4**	**5**	**6**
μ_{G_1}	0.8	0.65	0.7	0.9	0.9	0.85
μ_{G_2}	0.75	0.72	0.69	0.8	0.8	0.9
μ_{G_3}	0.7	0.68	0.65	0.75	0.8	0.8
μ_{G_4}	0.81	0.8	0.6	0.85	0.6	0.95

where we have six alternatives to select from and four fuzzy goals or constraints.

References

Bellman, R. and Zadeh, L. A., Decision making in a fuzzy environment, *Manage. Sci.*, 17(4), 141–164, 1970.

Bose, N. K. and Liang, P., *Neural Network Fundamentals with Graphs, Algorithms, and Applications*, McGraw-Hill, New York, 1996.

DeGroot, M. H., *Optimal Stochastic Decisions*, McGraw-Hill, New York, 1970.

Hecht-Nielsen, R., *Neurocomputing*, Addison-Wesley, Reading, MA, 1990.

Luenberger, D., *Introduction to Dynamic Systems*, John Wiley & Sons, New York, 1979.

Milton, J. S. and Arnold, J. C., *Introduction to Probability and Statistics*, 3rd ed., McGraw-Hill, New York, 1995.

Ross, Sh. M., *Introduction to Probability and Statistics for Engineers and Scientists*, John Wiley & Sons, New York, 1987.

Rubinstein, R. Y., *Simulation and the Monte Carlo Method*, John Wiley & Sons, New York, 1981.

Szidarovszky, F. and Bahill, A. T., *Linear Systems Theory*, 2nd ed., CRC Press, Boca Raton, FL, 1992.

Szidarovszky, F. and Yakowitz, S., *Principles and Procedures of Numerical Analysis*, Plenum, New York, 1978.

Szidarovszky, F., Geshon, M., and Duckstein, L., *Techniques for Multi-Objective Decision Making in Systems Management*, Elsevier, Amsterdam, The Netherlands, 1986.

Wets, R. J.-B., Stochastic programming: solution techniques and approximation schemes, in *Mathematical Programming: The State of the Art*, Bachem, A., Groetschel M., and Korte, B., Eds., Springer-Verlag, Berlin, 1983, pp. 506–603.

Yakowitz, S., Dynamic programming applications in water resources, *Water Resources Res.*, 18(4), 673–696, 1982.

4 Water Resources Economics

4.1 INTRODUCTION

Economics consists of a set of analytical tools that can be used for balancing competing objectives by allocating the scarce resources. *Engineering economy* is concerned with the economic aspects of engineering; it involves the systematic evaluation of costs and benefits of technical projects. *Environmental economics* is the application of the principles of economics to the study of how environmental resources are developed and managed.

The entire process of planning, design, construction, operation, and maintenance of water resources systems entails many important and complex decisions. Besides technological and environmental considerations, economic principles play a significant role in making these decisions. The principles of engineering economics guide the decision makers in selecting the best planning and/or operational decisions.

This chapter discusses the principles of engineering economics and their application to water resources planning and operation. Major objectives of economic analysis for water resources development projects are also discussed, and various methods for incorporating the money–time relationship are presented.

4.2 BENEFITS AND COSTS

Water resources development projects have two sides: on the one side, they create value and on the other side they encounter costs. The value side of the analysis is based on the fact that individuals have *preferences* for goods and services. The value of a good to a person is what that person is willing to pay for that good. Thus, the value of a good or service is tied to *willingness to pay.* The *marginal willingness to pay* is what each person is willing to pay for an additional unit of a good or service.

The costs associated with different economic activities can be classified as *fixed* and *variable*. Fixed costs are not affected by the range of operation or activity level. General management and administrative salaries and taxes on facilities are some examples of fixed costs. Variable costs are those associated with the quantity of output or other measures of activity level. An *incremental cost* or *incremental revenue* is the additional cost or revenue that results from increasing the output of a system by one unit.

Example 4.1

A contractor has the choice of two sites from which to deliver gravel to a construction site. The cost of transferring gravel to the site of an earthfill dam is estimated to be \$1.15/m^3/km. The costs associated with the two sites can be summarized as follows:

Cost	Site A	Site B
Distance from construction site	6 km	4.3 km
Monthly rental of site	\$1000	\$5000
Cost to set up and remove equipment	\$15,000	\$25,000
Transfer cost	\$1.15/m^3/km	\$1.15/m^3/km

The contract requires 50,000 km^3 of gravel to be delivered by the contractor, and it is estimated that 4 months will be required for the job. Compare the two sites in terms of their fixed, variable, and total costs. For the selected site, how many cubic meters of material must the contractor deliver before starting to make a profit if he is paid \$8.05/m^3 delivered to the construction site?

Solution: As shown in the following table, the rent and plant setup are fixed costs, whereas the transfer cost is a variable cost.

Costs	Fixed	Variable	Site A	Site B
Rent (4 months)	√		\$4000	\$20,000
Plant setup	√		\$15,000	\$25,000
Material transfer cost		√	6(50,000)(\$1.15) = \$345,000	4.3(50,000)(\$1.15) = \$247,250
Total			\$364,000	\$292,250

As can be seen in the table, site B, which has higher fixed costs, has the smaller total cost. The contractor begins to make a profit at the point where total revenue equals the total cost of the material delivered. For site B, it can be written that:

4.3(\$1.15) = \$4.95 in variable cost per cubic meter delivered.
Total cost = Total revenue.
\$45,000 + \$4.95X = \$8.05X.
X = 14,516 m^3 delivered.

Therefore, by selecting site B, the contractor will begin to make a profit on the job after delivering 14,516 m^3 of material.

Another classification of costs is *private* vs. *social costs.* The private costs are the costs experienced by the decision makers directly, but the social costs include all costs of an action, no matter who experiences them (Field, 1997). The difference between social and private costs is the *external cost* or *environmental cost*. These costs are called "external" because, although they are real costs to some members of society, the firms or agencies that are making decisions and taking actions do not

normally consider them when they make decisions. Water resources, which are *open access resources*, are important sources of external or environmental costs.

Example 4.2

Suppose three similar industries are situated on a lake. The industries use the water of the lake and discharge sewage back into the lake. Before using the water, the industries treat the water, which costs about $40,000 per year for each industrial unit. Suppose a new industrial unit is planning to start operating on the lake. Because of the untreated sewage being added to the lake by this unit, each of the existing industries will have to spend an additional $20,000 each year for water treatment. How much is the external cost of the new industrial unit operation?

Solution: Because each industrial unit will spend $20,000 per year for the additional treatment, the external cost is $60,000 ($20,000 × 3).

A similar definition is used by economists for *external benefits*. An external benefit is a benefit enjoyed by somebody who is outside of, or external to, decisions being made about consuming resources or using services. For example, consider a private power company that has constructed a dam for producing hydropower. Even though the dam produces downstream benefits such as mitigating both floods and low flows, the power company is not reimbursed for them; therefore, these benefits are classified as external benefits.

Although an economist seeks to evaluate all consequences of an action in commensurate monetary units, many values cannot be measured within such a framework. For example, many actions have direct physical effects on humans such as loss of health or life, emotional effects such as loss of prestige or personal integrity, and psychological effects such as environmental changes. These effects cannot be measured in monetary units; however, various economists have devoted a great deal of energy to quantifying these positive and negative effects within a monetary framework. A value that cannot be expressed in monetary terms is called *intangible* or *irreducible* (James and Lee, 1971).

4.3 MONEY–TIME RELATIONSHIP

Most engineering activities involve the commitment of capital for extended periods; therefore, the effects of time should be considered. The same amount of money spent or received at different times has different values because opportunities are available to invest the money in various enterprises to produce a return over a period.

Generally, the future value of a particular amount of money will be larger than the existing amount today. Financial institutions are willing to pay interest on deposits because they can lend the money to investors. If an amount of money is deposited in a bank, interest accrues at regular time intervals. These time intervals are called *interest periods*. The *interest rate* can be defined as the rate at which money increases in value from the present to the future. The *discount rate* refers to the rate by which the value of money is discounted from the future to the present. The interest accrued in a single interest period is called *simple interest*. If the earned

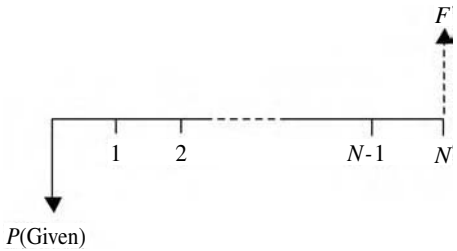


FIGURE 4.1 General cash flow diagram relating present equivalent and future equivalent of single payments.

interest is not withdrawn at the end of an interest period and is redeposited with the original investment in the next interest period, the interest thus accrued is referred to as *compound interest*.

The following notation is used in this section:

- i: Interest rate per interest period
- N: Number of compounding periods
- P: Present sum of money; the equivalent value of one or more cash flows at a reference point in time called the present
- F: Future sum of money; the equivalent value of one or more cash flows at a reference point in time called the future
- A: End-of-period cash flows (or equivalent end-of-period values) in a uniform series continuing for a specific number of periods, starting at the end of the first period and continuing through the last period
- G: End-of-period uniform gradient cash flows

4.3.1 Future Value and Present Value of Money

Figure 4.1 is a general cash flow diagram relating present, P, and future, F, single sums separated by N periods with an interest rate of i percent per period. As can be seen in this figure, P dollars is invested at the beginning of period one. This amount will grow to $P + Pi = P(1 + i)$ by the end of period one. Therefore, the future amount after N periods is equal to:

$$F = P(1+i)^N \tag{4.1}$$

The quantity $(1 + i)^N$ is the *single-payment compound amount factor*. The $(F/P, i\%, N)$ functional symbol is also used in this book for $(1 + i)^N$. For finding P when F is given, the following relation can be used:

$$P = F\left(\frac{1}{1+i}\right)^N = F(1+i)^{-N} = F\left(P/F, i\%, N\right) \tag{4.2}$$

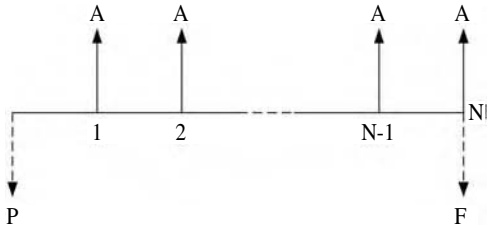


FIGURE 4.2 General cash flow diagram relating series of cash flows to present and future equivalent values.

The quantity $(1 + i)^{-N}$ is the *single-payment present worth factor.*

4.3.2 Future Value of Money and End-of-Period Cash Flows

Consider the cash flow diagram shown in Figure 4.2. We can treat each of cash flows separately and find their future values as follows:

$$F = A\left[(1+i)^{N-1} + (1+i)^{N-2} + \cdots + (1+i)^{1} + (1+i)^{0}\right]$$

which can be reduced to:

$$F = A\left[\frac{(1+i)^{N}-1}{i}\right] = A\left(F/A, i\%, N\right) \tag{4.3}$$

The quantity $(F/A, i\%, N)$ is the *uniform series compound amount factor.* For finding A, when F is given, the following relation can be used:

$$A = F\left[\frac{i}{(1+i)^{N}-1}\right] = F\left(A/F, i\%, N\right) \tag{4.4}$$

$(A/F, i\%, N)$ is called *sinking fund factor* or *capital recovery factor.* This factor indicates the number of dollars one can withdraw in equal amounts at the end of each of N periods if \$1 is initially deposited at $i\%$ interest.

4.3.3 Present Value of Money and End-of-Period Cash Flows

Using Eqs. (4.1) and (4.3) we obtain (see Figure 4.2):

$$P = A\left[\frac{(1+i)^N - 1}{i(1+i)^N}\right] = A\left(P/A, i\%, N\right) \tag{4.5}$$

(P/A, $i\%$, N) is the *uniform series present worth factor*. To find A when P is given, the following relation can be used:

$$A = P\left[\frac{i(1+i)^N}{(1+i)^N - 1}\right] = P\left(A/P, i\%, N\right) \tag{4.6}$$

The quantity (A/P, $i\%$, N) is the *capital recovery factor*.

4.3.4 Future Value of Money and End-of-Period Uniform Gradient Cash Flows

Figure 4.3 shows a cash flow diagram of a sequence of end-of-period cash flows increasing by a constant amount, G, in each period. The future equivalent of sequence of cash flows shown in Figure 4.3 can be computed as follows:

$$F = G\left(F/A, i\%, N-1\right) + G\left(F/A, i\%, N-2\right) + \cdots + G\left(F/A, i\%, 2\right) + G\left(F/A, i\%, 1\right)$$

$$F = G\left[\frac{(1+i)^{N-1} - 1}{i} + \frac{(1+i)^{N-2} - 1}{i} + \cdots + \frac{(1+i)^2 - 1}{i} + \frac{(1+i)^1 - 1}{i}\right]$$

which reduces to:

$$F = \frac{G}{i}\left[\sum_{j=0}^{N-1}(1+i)^j\right] - \frac{NG}{i}$$

$$F = \frac{G}{i}\left(F/A, i\%, N\right) - \frac{NG}{i} \tag{4.7}$$

4.3.5 Present Value of Money and End-of-Period Uniform Gradient Cash Flows

The following relation can be used for estimating the present equivalent value of cash flows shown in Figure 4.3:

$$P = \frac{G}{(1+i)^2} + \frac{2G}{(1+i)^3} + \cdots + \frac{(N-2)G}{(1+i)^{N-1}} + \frac{(N-1)G}{(1+i)^N}$$

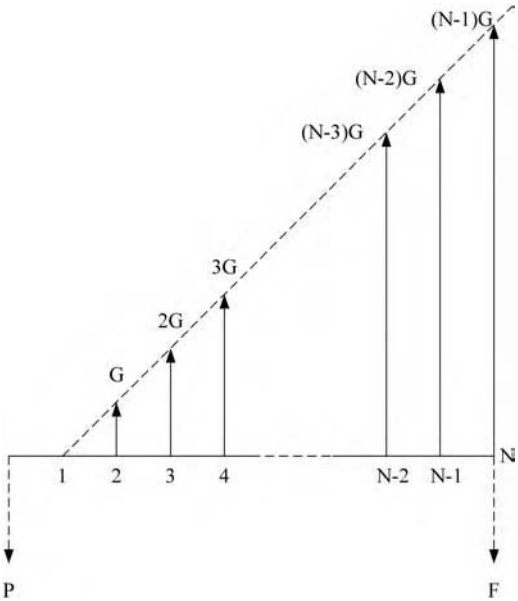


FIGURE 4.3 General cash flow diagram relating the uniform gradient of increasing cash flow to future and present equivalent values.

which reduces to:

$$P = G\left[\frac{(1+i)^N - 1 - Ni}{i^2(1+i)^N}\right] = G\left(P/G, i\%, N\right) \tag{4.8}$$

The factor (*P*/*G*, *i*%, *N*) is the *gradient to present equivalent conversion factor* or the *discount gradient factor.*

4.3.6 End-of-Period Cash Flows and End-of-Period Uniform Gradient Cash Flows

From Eq. (4.6), it can be written that:

$$A = P\left[\frac{i(1+i)^N}{(1+i)^N - 1}\right] = G\left[\frac{(1+i)^N - 1 - Ni}{i(1+i)^N - i}\right] = G\left(A/G, i\%, N\right) \tag{4.9}$$

The factor (*A*/*G*, *i*%, *N*) is the *uniform series gradient factor.* Table 4.1 provides a summary of previous sections about discrete compounding interest factors.

TABLE 4.1
Relations between General Cash Flow Elements Using Discrete Compounding Interest Factors

Relations	Interest Factor	Factor Name
$F = (1 + i)^N$	$(F/P,i\%,N)$	Single payment compound amount
$P = \dfrac{1}{(1+i)^N}$	$(P/F,i\%,N)$	Single payment present worth
$F = A \cdot \dfrac{(1+i)^N - 1}{i}$	$(F/A,i\%,N)$	Uniform series compound amount
$P = A \cdot \dfrac{(1+i)^N - 1}{i(1+i)^N}$	$(P/A,i\%,N)$	Uniform series present worth
$A = F \cdot \dfrac{i}{(1+i)^N - 1}$	$(A/F,i\%,N)$	Sinking fund
$A = P \cdot \dfrac{i(1+i)^N}{(1+i)^N - 1}$	$(A/P,i\%,N)$	Capital recovery
$P = G \cdot \dfrac{(1+i)^N - 1 - Ni}{i^2(1+i)^N}$	$(P/G,i\%,N)$	Discount gradient
$A = G \cdot \dfrac{(1+i)^N - 1 - Ni}{i(1+i)^N - i}$	$(A/G,i\%,N)$	Uniform series gradient

Example 4.3

Suppose you deposit \$6000 today, \$3000 after 2 years, and \$4000 after 5 years into a bank account paying 5% interest per year. How much money will you have in your account after 10 years?

Solution: The general cash flow diagram is shown in Figure 4.4. The money in the account after 10 years can be estimated as summation of the future values of single payments as follows:

$$F = 6000\left(F/P,5\%,10\right) + 3000\left(F/P,5\%,8\right) + 4000\left(F/P,5\%,5\right)$$
$$= \$6000(1.6289) + \$3000(1.4775) + \$4000(1.2763) = \$19311.1$$

Example 4.4

A water resources project produces benefits that amount to \$10,000 in year 1 and increase on a uniform gradient to \$100,000 in year 10, after which the benefits increase at a uniform gradient of \$5000 per year to \$175,000 in year 25. The value of the benefits then remains constant at \$175,000 each year until the end of the

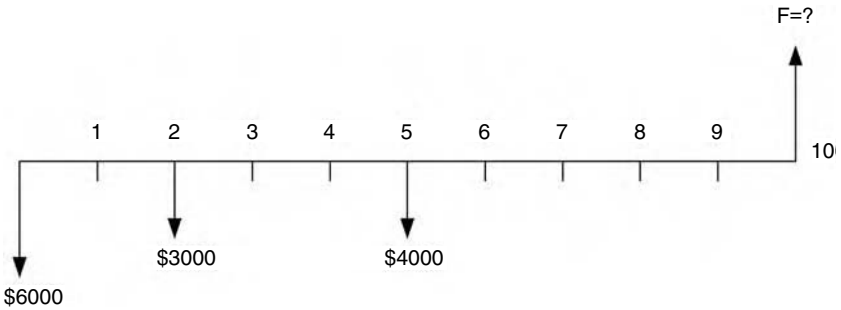


FIGURE 4.4 Cash flow diagram for Example 4.3.

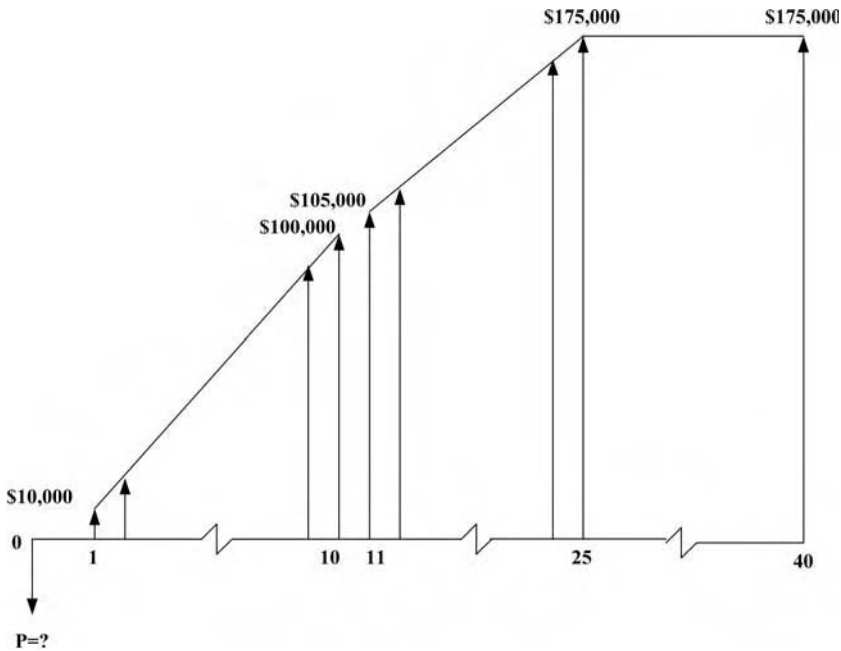


FIGURE 4.5 Cash flow diagram for Example 4.4.

project life, which is 40 years. What is the present worth of these benefits at a 4% interest rate? Figure 4.5 is a cash flow diagram for this example.

Solution:

1. Present equivalent value of benefits for years 1 through 10:

$$10,000\left(\frac{P}{G},4\%,10\right)+10,000\left(\frac{P}{A},4\%,10\right)=10,000\times 41.9919=\$419,919$$

2. Present equivalent value of benefits for years 11 through 25 is:

$$105,000\left(\frac{P}{A},4\%,15\right)\left(\frac{P}{F},4\%,10\right)+5,000\left(\frac{P}{G},4\%,15\right)\left(\frac{P}{F},4\%,10\right)$$
$$=105,000\times11.11839\times0.67556+5,000\times69.735\times0.67556=1,024,220$$

3. Present equivalent value of benefits for years 26 through 40 is:

$$175,000\left(\frac{P}{A},4\%,15\right)\left(\frac{P}{F},4\%,25\right)=175,000\times11.11838\times0.37512=729,877$$

The total present worth is the sum of these three values, which is equal to $2,174,016.

More details about interest rates that vary with time and interest formulas for continuous compounding can be found in DeGarmo et al. (1997) and Au and Au (1992).

4.4 ECONOMIC ANALYSIS OF MULTIPLE ALTERNATIVES

The main goal of economic analysis is to find out whether a capital investment and the costs associated with the project within the project lifetime can be recovered by revenues. In addition, it is important to find out whether the investment is sufficiently attractive considering the risks involved and potential alternative uses. The five methods explained briefly in this section are present worth, future worth, annual worth, internal rate of return, and external rate of return. When comparing alternatives, the selection of one alternative might exclude the others; these alternatives are *mutually exclusive*. Different projects usually have different useful lives, which should be reflected in the economic analysis.

4.4.1 Present Worth Method

In this method, the alternative that has the largest present worth (PW) of the discounted sum of benefits minus costs over its life is selected:

$$PW=\sum_{t=1}^{N}\left(B_t-C_t\right)\left(\frac{P}{F},i\%,t\right) \tag{4.10}$$

where:

C_t is the cost of each alternative in year t.
B_t is the benefit of each alternative in year t.
N is the study period in years or the planning horizon.
i is the interest rate.

By using this method to evaluate investment alternatives, the one with the greatest positive equivalent worth or least negative equivalent worth is selected. (Cost alternatives are those that will generate negative cash flows, except for a possible positive cash flow element derived from disposal of assets at the end of the useful life of the project; see DeGarmo et al. 1997).

When using this method, the following items should be considered:

- The same discount rate should be used for all alternatives.
- The same time base should be used for estimating the present worth of alternatives.
- The same study period (planning horizon) should be used.

Selection of the study period is an important step in comparing projects with different life spans. The study period is the selected time over which mutually exclusive alternatives are compared. The *repeatability assumption* can be used when the project lives are different.

Example 4.5

Consider two water resources development projects that are being proposed to supply the water demand of an industrial complex. Table 4.2 shows the characteristics of these projects. Find the best alternative when the minimum attractive rate of return (MARR) = 5%.

Solution: In this example, by using the repeatability assumption the study period is set at 20 years, which is the least common multiple of lives, and we consider two cycles of alternative B. By applying the present worth method, it can be written that:

$$A = -\$40,000 - \$160\left(\frac{P}{A}, 5\%, 20\right) + \$4,000 \times \left(\frac{P}{A}, 5\%, 20\right)$$

The present worth of alternative A is:

$$A = -\$40,000 - \$1,994 + \$49,845 = \$7,855$$

TABLE 4.2
Data for Example 4.5

Costs and Benefits	Project A	Project B
Construction costs	$40,000	$25,000
Operation and maintenance costs	$160	$100
Economic life	20 years	10 years
Annual benefits	$4000	$4000

The present worth of alternative B is:

$$\begin{aligned} B = & -\$25,000 - \$100\left(\frac{P}{A},5\%,10\right) + \$4,000\left(\frac{P}{A},5\%,10\right) \\ & -\$25,000\left(\frac{P}{F},5\%,10\right) - \$100\left(\frac{P}{A},5\%,10\right)\left(\frac{P}{F},5\%,10\right) \\ & +\$4,000\left(\frac{P}{A},5\%,10\right)\left(\frac{P}{F},5\%,10\right) = \$8,257 \end{aligned}$$

Therefore, project B should be chosen.

4.4.2 Future Worth Method

In this method, all benefits and costs of different alternatives are converted into their future worth figures. Then, the alternative with the greatest future worth or the least negative future worth is selected. When using this method, the following items should be considered:

- The same discount rate should be considered for all alternatives.
- The same time base should be used for estimating the future worth of alternatives.
- The same study period should be used for all alternatives.

When selecting the study period, two options may be considered. One option is to use the repeatability assumption, explained in the previous section. The second option is to use the *co-terminated assumption*. Based on this assumption, if an alternative has a useful life shorter than the study period, the estimated annual cost of the activities might be used during the remaining years. Similarly, if the useful life of an alternative is longer than the study period, a reestimated market value is normally used as the terminal cash flow at the end of the co-terminated life of the project (DeGarmo et al., 1997).

Example 4.6

Referring back to Example 4.5, assume that you want to invest your money in one of two alternatives. Consider that the study period is 20 years. Use the repeatability assumption and select the best alternative.

Solution: The future worth of alternative A can be estimated as follows:

$$\begin{aligned} \text{FW of } A &= -\$40,000\left(\frac{F}{P},5\%,20\right) - \$160\left(\frac{F}{A},5\%,20\right) + \$4,000\left(\frac{F}{A},5\%,20\right) \\ &= -\$40,000 \times 2.6533 - \$160 \times 33.0660 + \$4,000 \times 33.0660 = \$20,841 \end{aligned}$$

For estimating the future worth of alternative B, it can be assumed that all cash flows will be reinvested until the end of the study period, so the future worth of the project can be estimated as follows:

$$\text{FW of } B = \left[-\$25,000\left(\frac{F}{P},5\%,10\right)+\$4,000\left(\frac{F}{A},5\%,10\right)\right]\left(\frac{F}{P},5\%,10\right)$$
$$-\$100\left(\frac{F}{A},5\%,20\right)-\$25,000\left(\frac{F}{P},5\%,10\right)+\$4,000\left(\frac{F}{A},5\%,10\right)$$
$$=\left[-\$25,000\times 1.6289+\$4,000\times 12.5779\right]\times 1.6289$$
$$-\$100\times 33.066-\$25,000\times 1.6289+\$4,000\times 12.5779=\$21902$$

Therefore, alternative B should be selected.

4.4.3 Annual Worth Method

In this method, the costs and benefits of all alternatives are converted to uniform annual figures, then the alternative with the greatest annual worth or the least negative worth, in the case of cost alternatives, is selected. When using this method, projects with different economic lives can be compared without considering the previously explained assumptions about the study period. It should be noted that when applying this method, the same discount rate should be used for all alternatives, just as for the present worth and future worth methods.

Example 4.7

Solve Example 4.5 by using the annual worth method.

Solution:

$$\text{AW of } A=-\$40,000\left(\frac{A}{P},5\%,20\right)-\$160+\$4,000=\$632$$

$$\text{Aw of } B=-\$25,000\left(\frac{A}{P},5\%,10\right)-\$100+\$4,000=\$663$$

Alternative B is selected because of its higher annual worth.

4.4.4 The Internal Rate of Return (IRR) Method

The internal rate of return (IRR) method is one of the most widely used methods for economic analysis. In this method, the IRR is defined as the discount rate that will set the net present value or the net future value of the cash flow profile equal to zero. In this method, the projects that have IRR less than the *minimum attractive rate of return* (MARR) will be rejected. The present worth and annual worth methods are usually used to find the IRR.

Example 4.8

Suppose you have invested \$1000 in a project. Based on the contract, you will receive benefits of \$500 and \$1500, 3 and 5 years later, respectively. How much is the IRR of the investment?

Solution: Using the present worth method, it can be written that:

$$-\$1,000 + \$500\,(P/F, i\%, 3) + \$1,500\,(P/F, i\%, 5) = 0$$
$$\Rightarrow \quad i = 16.90\%$$

So, if the MARR is less than 16.90%, the investment will not be rejected.

In this method, it is assumed that the recovered funds are reinvested at IRR. In many real situations, it is not possible to reinvest the money at IRR rather than at MARR. In such cases, the *external rate of return* (ERR) method should be used.

4.4.5 The External Rate of Return Method

As mentioned in the previous section, it is sometimes impossible to reinvest the recovered money at the internal rate of return. The external rate of return method takes into account the interest rate, e, that is external to a project and at which net cash flows generated by the project over its life can be reinvested. This method can be summarized in the following steps:

1. The present value of all net cash flows is estimated using the e percentage discount rate.
2. The future value of all net cash flows is estimated using the e percentage interest rate.
3. The external rate of return is estimated by setting these two quantities equal.

Example 4.9

Consider the cash flow diagram shown in Figure 4.6. If the external rate of return is considered 15% and the MARR is equal to 20%, is the project rejected?

Solution: Based on the ERR method, it can be written that:

$$ABS\left[\left(-\$1,000 - \$500\left(P/F, 15\%, 1\right)\right) \times \left(F/P, i'\%, 6\right)\right]$$
$$= \left(\$600 - \$100\right) \times \left(F/A, 15\%, 5\right) \quad \Rightarrow \quad i'$$
$$i' = 15.3\%$$

Because i' is less than MARR, the project is rejected.

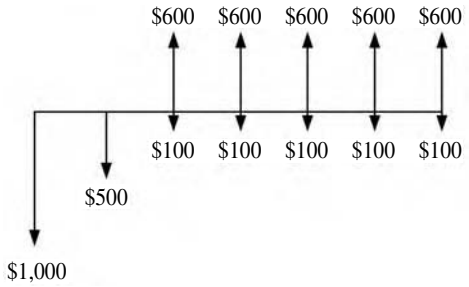


FIGURE 4.6 Cash flow diagram for Example 4.9.

4.4.6 Comparing Alternatives Using the Rate of Return Method

When comparing different alternatives using the rate of return method, the same study period should be considered. Projects that have a rate of return lower than the MARR would be rejected. The more costly alternative would be selected only if the incremental rate of return exceeds the MARR; otherwise, the least costly alternative is selected.

Example 4.10

Two different types of pumps (A, B) can be used in a pumping station. Table 4.3 shows the costs and benefits of each of the pumps. Compare the two pumps using a MARR = 10%.

Solution: The study period is 12 years. By using the repeatability assumption, it can be written that:

Present worth of costs of A = \$200 + (\$200 – \$50)(*P*/*F*, 10%, 6) – \$50 (*P*/*F*, 10%, 12) = \$269
Present worth of benefits of A = \$95(*P*/*A*, 10%, 12) = \$647
Present worth of costs of B = \$700 – \$150(*P*/*F*, 10%, 12) = \$652
Present worth of benefits of B = \$120(*P*/*A*, 10%, 12) = \$818

TABLE 4.3
Data Used in Example 4.10

Economic Characteristics	Pump A	Pump B
Initial investment	\$200	\$700
Annual benefits	\$95	\$120
Salvage value	\$50	\$150
Useful life (year)	6	12

The graphical method can be used to select one of the pumps. In this method, a line is drawn between the two points representing the present costs vs. present benefits

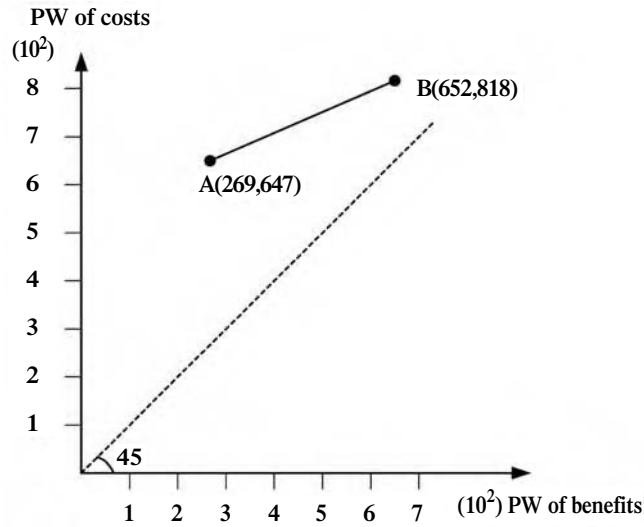


FIGURE 4.7 Comparing the two projects in Example 4.10 by the rate of return method.

for A and B. The slope of this line is compared with a 45° line drawn on the graph. As can be seen in Figure 4.7, the present worth of costs and the present worth of benefits are equal for points on the line having a 45° slope. When the slope of the line connecting two projects is less (more) than 45°, the project with the lower (higher) initial cost would be selected. As shown in Figure 4.7, pump A, which has a lower initial cost, should be selected.

4.5 ECONOMIC EVALUATION OF PROJECTS USING THE BENEFIT–COST RATIO METHOD

The benefit-cost ratio method has been widely used in the economic analysis of water resources projects. This method is based on the calculation of the ratio of benefits to costs. The time value of money should be considered; therefore, this ratio should be estimated based on the equivalent worth of discounted benefits and costs. For this purpose, the annual worth, present worth, or future worth of benefits and costs can be used. One of the basic formulations for the benefit–cost ratio method is as follows:

$$B/C \text{ factor} = \frac{B}{I + C} \tag{4.11}$$

where:

B is the net equivalent benefits.
C is the net equivalent annual costs, usually classified as operation and maintenance costs.
I is the initial investment.

For any project to remain under consideration, its benefit–cost ratio should exceed one. Therefore,

$$\frac{B}{I+C} > 1$$

Then,

$$B-(I+C) > 0$$

Therefore, the benefit–cost criterion will eliminate all those projects for which equivalent amounts are less than zero. Another formulation for this method, which reflects the net gain expected per dollar invested, is as follows:

$$B/C \text{ factor} = \frac{B-C}{I} \tag{4.12}$$

In order to consider the salvage value associated with an investment, the following relation can also be used in applying the benefit–cost method:

$$B/C \text{ factor} = \frac{B-C}{I-S} \tag{4.13}$$

where S is the salvage value of the investment.

Example 4.11

Referring back to Example 4.10, assume that the annual costs of operation and maintenance for pumps A and B are equal to \$50 and \$60, respectively. Compare the two pumps using the benefit–cost ratio method.

Solution: Considering the salvage value of the pumps, Eq. (4.13) can be used. Because the useful lives of the projects are different, it is easier to use the annual worth method, as follows:

$$\text{Pump A:} \quad B/C = \frac{95-50}{200(A/P,10\%,6)-50(A/F,10\%,6)}$$

$$= \frac{45}{200\times 0.2296 - 50\times 0.1296} = 1.14$$

$$\text{Pump B:} \quad B/C = \frac{120-60}{700(A/P,10\%,12)-150(A/F,10\%,12)}$$

$$= \frac{60}{700\times 0.1468 - 150\times 0.0468} = 0.63$$

It is apparent that the benefit–cost ratio for pump B is less than one; therefore, pump A should be selected.

In order to compare multiple mutually exclusive alternatives by the benefit–cost ratio method, the principle of incremental analysis should be applied. For this purpose, the following steps should be taken:

1. Arrange the alternatives by increasing order of their equivalent worth of costs (denominators, or $I + C$, I, and $I - S$ in Eqs. (4.11), (4.12), and (4.13), respectively).
2. Consider the alternative with the denominator of least value as the baseline.
3. Estimate an incremental B/C ratio ($\Delta B/\Delta C$) based on the difference between the equivalent worth of benefits and costs of the second least cost alternative and the baseline.
4. If the ratio is equal to or greater than one, the higher equivalent cost alternative becomes the new baseline; otherwise, the last baseline alternative is maintained.
5. Repeat steps (3) and (4) until the last alternative has been compared.

Example 4.12

Three flood control projects have been proposed for a river. Each of the projects has a useful life of 25 years and the interest rate is 5% per year. The benefits and costs of the projects are listed in Table 4.4. Compare the alternatives using the benefit–cost ratio method.

Solution: Present worth of costs and benefits for the projects can be estimated as follows:

Present worth of costs of A = \$8,000,000 + \$700,000 (P/A, 5%, 25) = \$17,865,730
Present worth of costs of B = \$9,500,000 + \$650,000 (P/A, 5%, 25) = \$18,661,035
Present worth of costs of C = \$11,000,000 + \$630,000 (P/A, 5%, 25) = \$19,879,157
Present worth of benefits of A = \$1,500,000 (P/A, 5%, 25) = \$21,140,850
Present worth of benefits of B = \$1,600,000 (P/A, 5%, 25) = \$22,550,240
Present worth of benefits of C = \$1,750,000 (P/A, 5%, 25) = \$24,664,325

TABLE 4.4
Benefits and Costs of Three Flood Control Alternatives in Example 4.12 (in \$millions)

Benefits and Costs	Project A	Project B	Project C
Capital investment	8	9.5	11
Annual operation and maintenance costs	0.7	0.65	0.63
Annual expected reduction in flood damage (benefits)	1.5	1.6	1.75

For all the projects the benefit–cost ratio is greater than one; therefore, all the projects are acceptable. For comparing projects A and B, it can be written that:

$$\Delta B / \Delta C(B-A) = \frac{\$22,550,240 - \$21,140,850}{\$18,661,035 - \$17,865,730} = 1.77$$

Because the estimated benefit–cost ratio is greater than one for each project, project B should be considered as the baseline because it has a higher present worth of costs than project A. For comparing projects B and C, it can be written that:

$$\Delta B / \Delta C(C-B) = \frac{\$24,664,325 - \$22,550,240}{\$19,879,157 - \$18,661,035} = 1.73$$

The estimated ratio is greater than one; therefore, project C is recommended.

4.6 PUBLIC VS. PRIVATE PROJECTS AND GOODS

Water resources projects are usually classified as *public projects*, which means they are financed and operated by federal, state, or local governmental organizations. These projects may be of any size but are usually very large. These projects are subject to the principals of engineering economics, but a number of special characteristics should be included in the economic evaluation of these projects. Some of the basic differences between public water resources projects and privately owned projects could be summarized as follows:

- **Purposes of the project and conflict issues.** Public water resources projects are usually *multipurpose*. For example, a reservoir might supply water and power demands and instream flows, in addition to controlling floods to prevent damage that may occur downstream of the reservoir in high-flow seasons. Some of these objectives are in conflict, which is a common issue. For example, holding enough water in the reservoir to control flooding is in conflict with water conservation schemes for supplying demands in low-flow seasons. Conflict issues in regard to water resources development projects are explained in the next several chapters of this book. The main purpose of water resources in public sector is to *minimize social costs*. Private projects usually provide specific goods and services that are not in conflict.
- **Stakeholder community.** While private projects are often related to a limited community of stakeholders, water resources projects usually affect a large number of people. Besides the direct impact of supplying water demand for users of the water resources system, other issues such as environmental impacts or controlling power network fluctuations also have broad effects on regional or national scales.
- **Project life.** The life of private projects is relatively short, while the life of public projects can be as long as 20 to 100 years.

- **Measurement of the project efficiency.** For private projects, the rate of return on capital is usually considered as a measure of efficiency, while for water resources development projects the efficiency is almost always very difficult to measure. This is mainly because of different short- and long-term effects of water resources projects that cannot be easily quantified in monetary units. In contrast, the benefits of private projects are usually monetary or relatively easy to equate to monetary terms.
- **Political issues.** Funding for public projects is often subject to a strong political influence, while this is not an important issue for most private projects. In addition to public and private projects, goods also can be classified as public and private. Two main characteristics of public goods are that they are *nonrival* and *nonexclusive* (Mansfield, 1994). For nonrival goods, the marginal cost of providing goods to an additional consumer is zero. This means that one person's use does not preclude enjoyment by the others. Agricultural and industrial uses of water are rivals in consumption, while the aesthetic and recreational values of a beautiful stream are nonrivals. Also, we consider these goods to be nonexclusive because people cannot be excluded from consuming the goods. Costs required to keep those not entitled from using the pubic good are *exclusion costs*. Exclusion costs for water resources are usually very high; therefore, we usually consider them to be nonexclusive.

4.7 ECONOMIC MODELS

Economic models relate the needs of consumers to the resources that can be used to satisfy these needs. These models include demand functions for goods and services, supply functions for resources, and production functions for goods and services. Each of these elements is briefly explained in the following sections.

4.7.1 Production Functions

Production functions refer to production of useful products by utilizing resources and certain technologies. For example, supplying water for irrigation might require specific structures for diversion and transferring water to agricultural lands. The production technologies are subject to change and advancements, which cannot be predicted accurately. For economic activities that significantly affect environment and natural resources, the tradeoff between producing goods and services and environmental quality is an important issue. This tradeoff is illustrated in Figure 4.8 by a *production possibility curve*. Considering the technological capabilities over a specific time period, each society may choose to locate itself at a specific point on this curve. Therefore, the issue becomes one primarily of *social choice* and depends on the value that people in a particular society place on the environment and economic outputs (Field, 1997). As can be seen in the figure, the current level of economic output (C_1) can be increased to C_2 only at the cost of a decrease in environmental quality from E_1 to E_2.

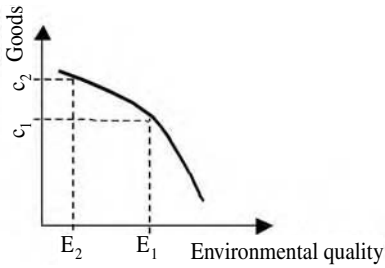


FIGURE 4.8 Production possibilities curve. (Adapted from Field, B. C., *Environmental Economics: An Introduction*, McGraw-Hill, New York, 1997.)

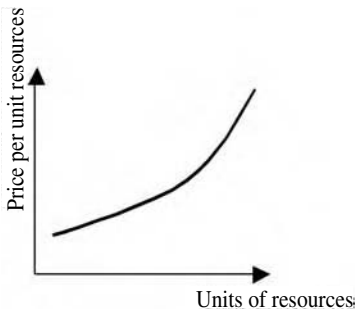


FIGURE 4.9 Resource-supply function.

4.7.2 Resource-Supply Functions

The resource-supply function may be defined as the quantity of resources made available per unit time as a function of the price per unit of resources. As can be seen in Figure 4.9, if the price for a given resource is high, the resource supplier will be willing to deliver a large quantity. In order to develop the water supply resources, development of well fields, interbasin water transfer projects, etc. could be considered, if the price of water is estimated accurately. The distribution of water resources development costs should be considered when making decisions about development alternatives and water allocation schemes. When a scarcity occurs, more water might be allocated to users who are capable of paying high prices for water, such as industries; however, political and legal issues may significantly affect this process (Viessman and Welty, 1985).

4.7.3 Demand Functions

Demand functions indicate the quantity of goods and services that the consumer would demand at any particular price. This is another way of looking at the marginal willingness to pay relationship. Figure 4.10 shows an example of demand functions for water resources. The price–quantity relation for domestic use is very steep. It means that even if the price of water significantly increases, people would still satisfy their needs for cooking, washing, and maintaining health standards. At the same

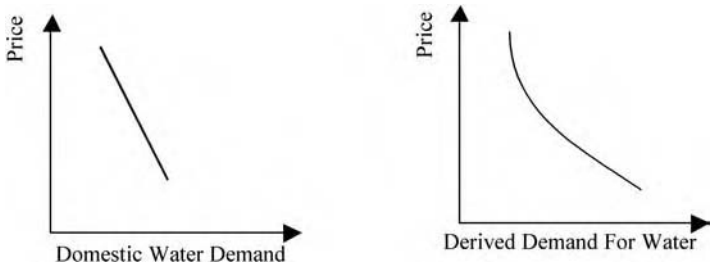


FIGURE 4.10 Examples of demand function. (Adapted from Viessman, W. and Welty, C., *Water Management: Technology and Institutions*, Harper & Row Publishers, New York, 1985.)

time, people are not going to use much more water than they really need no matter how inexpensive the water is. The demands for products derived from water are more sensitive to price changes.

The demand functions that engineers and planners encounter are more complex because of a variety of factors that affect the demand functions. The demand functions can be formulated as follows (Mays, 1996):

$$Q_W = Q_W\left(P_W, P_a, P, Y, Z\right) \tag{4.14}$$

Where Q_W refers to the consumers' water demands in a specific period of time, P_W is the price of water, P_a is the price of an alternative water resource, P is an average price index representing all other goods and services, Y is the consumer income, and Z is a vector representing other factors such as climate and consumer preferences.

The responsiveness of consumer demands to variations in each of the above-mentioned variables is an important concept in economic analysis and is referred to as *demand elasticity.* One of the most important aspects of demand elasticity is *price elasticity*, which can be defined as the percentage change in quantity consumed if price is changed 1%. Almost all estimates of long-run price elasticity of domestic water demand in the United States seem to fall between –0.3 and –0.7, meaning that other factors being held constant, a 10% increase in price would lead to a 3 to 7% decrease in the amount purchased (Mays, 1996).

4.8 EFFICIENCY AND EQUITY IN ECONOMIC ANALYSIS

A goal of economic analysis is to understand human behavior when facing scarcity to determine how to manage scarcities of resources and products. In the case of water resources management problems, allocation of water resources among competing users is the main issue. The two main effects of allocating scarce sources are classified as *efficiency* and *distribution* (*equity*). Economic efficiency is an index for measuring whether the allocation of resources among competing users has maximized profits or welfare level. For economic activities that use natural resources,

TABLE 4.5
Distribution of Costs and Benefits in Example 4.13

Benefits and Costs	Group 1	Group 2	Group 3
	($/person/year)		
Decrease in costs of water supply (benefits)	60	80	120
Costs of the project	40	60	80
Difference	20	20	40

such as water, or affect these resources by polluting wastes, efficiency should be measured by incorporating social costs.

Equity effects refer to allocation of costs and benefits to individuals and different groups in the society. Distribution of costs and benefits based on equity can follow a *vertical* or *horizontal distribution*. If the benefits of a water resources development project allocated to different groups of people with different levels of income are equal, then the distribution is horizontal. In vertical distribution, the costs and benefits are allocated to different groups of people proportional to the level of their income. In the vertical distribution of costs and benefits, if more profits are allocated to groups with higher levels of income, the program is *regressive*. In contrast, *progressive programs* allocate more profit to groups with lower levels of income.

Example 4.13

A water supply project is implemented in a region. Three groups of people with the same levels of income will use this project. Table 4.5 shows the distribution of costs and benefits among these groups. Is the distribution of costs and benefits vertical or horizontal?

Solution: As shown in Table 4.5, for groups 1 and 2 the difference between benefits and costs is equal; therefore, a horizontal distribution has been applied to groups 1 and 2 but the difference between costs and benefits for the third group is twice that of the other two groups. In the next example, the effects of vertical distribution are explained.

Example 4.14

Three water supply projects are implemented in a region. The water demand of people in a big city, a small city, and a village will be supplied by these projects. The levels of income and distributions of costs and benefits for these projects are presented in Table 4.6. Discuss the distributions of costs and benefits and determine which projects are progressive and regressive.

Solution: As shown in Table 4.6 for the first program, the difference between costs and benefits is 1% of income for the people in both cities and in the village; therefore, the distribution of costs and benefits is vertical in this program. The second program is a regressive program, because the profits allocated to the groups with higher levels of income are higher. The third program is progressive, because it has

TABLE 4.6
Distribution of Costs and Benefits in Three Water Supply Projects in Example 4.14

Benefits and Costs	Village (Pop. 5000)	Small City (Pop. 20,000)	Big City (Pop. 50,000)
	($/person/year)		
Water supply program 1			
Decrease in water supply from other sources (benefits)	150 (3)	300 (1.5)	600 (1.2)
Costs of project implementation ($)	100 (2)	100 (0.5)	100 (0.2)
Difference	50 (1)	200 (1)	500 (1)
Water supply program 2			
Decrease in water supply from other sources (benefits)	150 (3)	1400 (7)	5500 (11)
Costs of project implementation ($)	100 (2)	800 (4)	3000 (6)
Difference	50 (1)	600 (3)	2500 (5)
Water supply program 3			
Decrease in water supply from other sources (benefits)	700 (14)	2200 (11)	3000 (6)
Costs of project implementation ($)	200 (4)	1000 (5)	1500 (3)
Difference	500 (10)	1200 (6)	1500 (3)

Note: Numbers in the parentheses are percents of income

allocated less profit to the people with higher income. It should be noted, however, that a project might be regressive or progressive overall, but the distributions of costs and benefits might be different. For example, the second program is regressive, but the costs have been distributed progressively.

In quantifying economic efficiency, the differences between public and private goods should also be considered. For example, consider two groups of consumers for a good or service. The demand curves for these groups are represented by D_A and D_B. The total demand for these groups when the good is a private good can be estimated by horizontally summing the demand curves for each group. The total demand curve (D) and the demand curves for two consumers are shown in Figure 4.11.

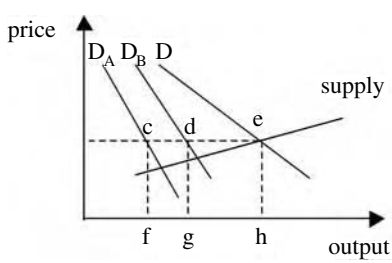


FIGURE 4.11 Efficient output of a private good.

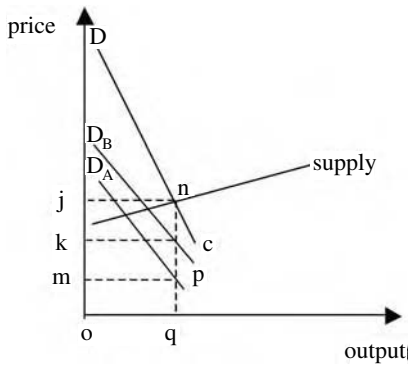


FIGURE 4.12 Efficient output of a public good.

The efficient output for this case is *h* because at this output the marginal benefit to each consumer of an extra unit of the good equals its marginal cost. Assuming that the marginal benefit can be measured as the maximum amount that each group will pay for the extra unit, then the marginal benefit for groups *A* and *B* would be *cf* and *dg*, respectively. The marginal cost of the extra unit is *eh* for the output rate equal to *h*. Because *cf* = *dg* = *eh*, the marginal benefit to each group equals the marginal cost (Mansfield, 1994).

Figure 4.12 shows a similar case for a public good. The total demand curve can be obtained by summing the demand curves for the two groups vertically. Because both groups of consumers consume the total amount of the good, the combined price paid by the two groups of consumers is the sum of the price paid by each group.

The efficient output of a public good occurs when the marginal social benefits equal the marginal social costs. For the output rate of *q*, the marginal social benefit from an extra unit of output is the vertical sum of *om* and *ok*, which equals *oj*. Because the marginal social cost of an extra unit of output is *qn* and because *oj* is equal to *qn*, *oq* is the optimal rate of supply for the public good.

4.9 PROBLEMS

4.1 What is the future amount that will be available in 4 years if $1000 is invested at 11% per year simple interest?

4.2 A company lends $10,000 at 10% interest rate for 10 years. At the end of year 10, the entire principal plus interest is invested at a 12% rate of interest for 10 years. How much will accumulate at the end of a 20-year period?

4.3 Draw a cash flow diagram for problem 4.2 from the viewpoint of the lending company.

4.4 A certain type of equipment is intended to be used in a project for which the economy study is based on a 25-year period; it is estimated that it will require replacement every 10 years. What salvage value should be used if the initial cost is $10,000? Use a 10% interest rate.

4.5 Following are three alternatives for supplying water to a community for 25 years, after which the economic lives and period of analysis terminate. Using the costs presented in the table and a 5% discount rate, compare the three projects using the present worth method.

Costs	Timing (Year)	Project A ($)	Project B ($)	Project C ($)
Initial investment	0	2,000,000	1,000,000	1,500,000
	10	0	1,000,000	1,200,000
	20	0	1,000,000	0
Operation and maintenance costs	1–10	7000	4000	6000
	10–20	8000	7000	8000
	20–25	9000	9000	9000

4.6 Compare the projects in problem 4.5 using the benefit–cost ratio method.

4.7 Compare the projects in problem 4.5 using the rate of return method.

4.8 Consider the following two mutually exclusive alternatives for a water diversion project:

	Project A	Project B
Capital investment	$200,000	$300,000
Annual revenues	$100,000	$140,000
Annual expenses	$44,000	$86,000
Useful life (years)	10	15

Assume the study period is 15 years and repeatability is applicable. Using the annual worth method, which project do you recommend?

4.9 Solve problem 4.8 using the present worth method.

4.10 The net cash flows are shown below for three alternatives for water resources development projects. The MARR is 12% per year and the study period is 10 years. Which alternative is economically preferred based on the external rate of return method? (External rate of return = MARR = 12% per year.)

End of Year	Project A ($)	Project B ($)	Project C ($)
0	171,200	126,400	143,600
1	14,800	24,200	20,000
?	?	?	?
?	?	?	?
?	?	?	?
10	14,800	24,200	20,000

4.11 Three water supply projects are implemented in a region. Water demand of people in a big city, a small city, and a village will be supplied through these projects. Level of income and distribution of costs and benefits for these projects are presented in the following table. Discuss the distribution of costs and benefits and determine which projects are progressive and regressive.

Benefits and Costs	Village (Pop. 1000)	Small City (Pop. 20,000)	Big City (Pop. 70,000)
	($/person/year)		
Water supply program 1			
Decrease in water supply from other sources (benefits)	30	1400	7700
Costs of project implementation	20	800	4200
Difference	10	600	3500
Water supply program 2			
Decrease in water supply from other sources (benefits)	140	2200	7700
Costs of project implementation	40	1000	4200
Difference	100	1200	3500
Water supply program 3			
Decrease in water supply from other sources (benefits)	30	300	840
Costs of project implementation	20	100	140
Difference	10	200	700

REFERENCES

Al-Futaisi, A. and Stedinger, J. R., Hydrologic and economic uncertainties and flood-risk project design, *ASCE J. Water Resources Plan. Manage.*, 125(6), 314–324, 1999.

Arnold, F. S., *Economic Analysis of Environmental Policy and Regulation*, John Wiley and Sons, New York, 1995.

Au, T. and Au, T. P., *Engineering Economics for Capital Investment Analysis*, Boston: Allyn and Bacon, 1983.

Creel, M. and Loomis, J., Recreation value of water to wetlands in the San Joaquin Valley: linked multinomial logic and count data trip frequency model, *Water Resources Res.*, 28(10), 1992.

DeGarmo, E. P., Sullivan, W. G., Bontadelli, J. A., and Wicks, E. M., *Engineering Economy*, Prentice-Hall, Upper Saddle River, NJ, 1997.

Feder, G. and Le Moigne, G., Managing water in sustainable manner, *Finance Dev.*, 31(2), June 1994.

Field, B. C., *Environmental Economics: An Introduction*, McGraw-Hill, New York, 1997.

Hartwick, J. H. and Olewiler, N. D., *The Economics of Natural Resources Use*, Addison-Wesley, Reading, MA, 1998.

Hoehn, J. P. and Krieger, D. J., Economic analysis of water services investments and tariffs in Cairo, Egypt, *ASCE J. Water Resources Plan. Manage.*, 126(6), 345–350, 2000.

James, L. D. and Lee, R. R., *Economics of Water Resources Planning*, McGraw-Hill, New York.

Kahn, J. R., *The Economic Approach to Environmental and Natural Resources*, 2nd ed., Dryden Press, Troy, MO, 1998.

Kizner, I. M., Entrepreneurial discovery and the competitive market process: an Austrian approach, *J. Econ. Lit.*, 35(1), 60–85, 1997.

Loomis, J. B., *Integrated Public Lands Management: Principles and Applications on National Forests, Parks, Wildlife Refuges, and BLM Lands*, Columbia University Press, New York, 1993.

Loomis, J. B. and Walsh, R. G., *Recreation Economic Decisions: Comparing Benefits and Costs*, Department of Agriculture and Resources Economics, Colorado State University, Fort Collins, CO, 1997.

Loomis, J. B., Environmental valuation techniques in water resources decision making, *ASCE J. Water Resources Plan. Manage.*, 126(6), 339–344, 2000.

Maksimovic, C. and Tejada-Guibert J. A., Eds., *Frontiers in Urban Water Management: Deadlock or Hope*, IWA, London, 2001.

Mansfield, E., *Managerial Economics: Theory, Application, and Cases*, W.W. Norton, 1996.

Mansfield, E., *Microeconomics: Theory and Applications*, W. W. Norton, New York, 1994.

Mays, L. W., *Water Resources Handbook*, McGraw-Hill, New York, 1996.

Pudney, S., *Modeling Individual Choice: The Econometrics of Corners, Kinks, and Holes*, Basel-Blackwell, Oxford, 1989.

Recio, B., Rubio, F., Lomban, J., and Ibanez, J., An econometric irrigated crop allocation model for analyzing the impact of water restriction policies, *J. Agric. Water Manage.*, 42, 47–63, 1999.

Sagoff, M., *The Economy of the Earth*, Cambridge University Press, Landon, 1988.

Shabman, L. and Stephenson, K., Environmental valuation and its economic critics, *ASCE J. Water Resources Plan. Manage.*, 126(6), 382–388, 2000.

Tietenberg, T., *Environmental and Natural Resources Economics*, Harper Collins, New York, 1992.

United Nations, *Proceedings of the Expert Group Meeting on Water Pricing*, Water Resources Series, No. 55, United Nations, New York, 1981.

USACE, *Guidelines for Risk and Uncertainty Analysis in Water Resources Planning*, Water Resources Support Center, Institute of Water Resources, U.S. Army Corps of Engineers, Fort Belvoir, VA, 1992a.

USACE, *Guidelines and Procedures for Risk and Uncertainty Analysis in Corps Civil Works Planning*, Water Resources Support Center, Institute of Water Resources, U.S. Army Corps of Engineers, Fort Belvoir, VA, 1992b.

USWRC, *Economics and Environmental Principles for Water and Related Land Resources Implementation Studies*, U.S. Water Resources Council, U.S. Government Printing Office, Washington, D.C., 1983.

Viessman, W. and Welty, C., *Water Management: Technology and Institutions*, Harper & Row Publishers, New York, 1985.

5 Time Series Analysis

5.1 INTRODUCTION

The concept of random variables has been used in the field of hydrology and water resources since the beginning of the 20th century, but in recent years a great deal of effort has been devoted to improve the following:

1. Understanding the stochastic nature of hydrologic variables such as streamflows
2. Modeling procedures
3. Developing new statistical models
4. Parameter estimation techniques
5. Model evaluation and fitness tests

Operation policies for water resources systems could not be applied in real-time decision making without forecasting the future state of the resources. For example, consider a reservoir that supplies water for various purposes; the amount of each scheduled release depends on the probable range of inflow to the reservoir. Because of a lack of adequate knowledge about physical processes in the hydrologic cycle, many investigators have expanded the application of statistical models to forecasting and generation of synthetic data. Synthetic data also help in incorporating uncertainties and probable extreme events.

New advancements in this area have been mainly focused on coupling physical characteristics of the water resources systems and the effects of large-scale climate signals on water resources and early prediction of rainfall and streamflow. This chapter discusses the basic principles of hydrologic time-series modeling and several types of statistical models. It should be noted that the details of statistical modeling are presented in many references, such as Salas et al. (1988) and Brockwell and Davis (1987). An introduction to these models and their application in water resources systems analysis are presented here.

5.2 STOCHASTIC PROCESSES

All natural physical processes are subject to variability. For example, rainfall intensity, flood magnitude, or low flows during severe droughts can vary widely. To study these variations and incorporate them in the planning and operation of water resources, *samples* of data are collected. A dataset consists of a series of measurements of a phenomenon, and the quantities measured are termed *variables*. To allow for the possible unpredictable nature of future observations, it is assumed that each observation is a realized value of a certain random variable.

The hydrologic cycle is composed of phenomena such as precipitation, runoff, infiltration, evaporation, evapotranspiration, and abstraction. *Characteristic variables*, which can also be referred to as *hydrologic variables*, have been defined to describe each of these phenomena. Depth or intensity of rainfall at different time steps of a rainstorm, monthly inflow discharge to a reservoir, or daily evaporation are some examples of hydrologic variables, which usually vary in time and space (Shahin et al., 1993). A *time series* is a sequence of values arranged in their order of occurrence in time.

A *process* is a mathematical description of the behavior of a phenomenon in one or more dimensions in space and/or time. Because all hydrologic phenomena change in space or time, they are *hydrologic processes*. If a process contains a random component, it is a *stochastic process*, which is a family of random variables, $\{x_t, t \in T\}$, defined on a probability space. Stochastic processes are subdivided into *stationary* and *nonstationary*. A stochastic process is stationary if the expected values of statistical descriptors do not change over time. If a time series is stationary, the series should be divided into a number of nonoverlapping subseries and the expected values of statistical descriptors of each series should be the same for each of the subseries.

Hydrologic variables are mostly nonstationary due to variations that are the result of natural and human activities. In hydrologic modeling, these variations can be classified as:

- *Trend* is a unidirectional gradual change (increasing or decreasing) in the average value of the variable. Changes in nature caused by human activities are the main reason for the over-several-years trends. The trends are usually smooth, and we should be able to represent it by a continuous and differentiable function of time. Trend, T_t, is usually considered to be deterministic, and it can be modeled by linear or polynomial functions, such as:

  Linear function: $T_t = a + bt$

  Polynomial function: $T_t = a + bt + ct^2 + \cdots + dt^m$

  Power functions: $T_t = a - br^t \;\; (0 < r < 1)$ and $T_t = 1/(a + br^t) \;\; (0 < r < 1)$

- *Jump* is a sudden positive or negative change in the observed values. Human activities and natural disruptions are the main reasons for jumps in hydrologic series.
- *Periodicity* represents cyclic variations in a hydrologic time series. These variations are repetitive over fixed intervals of time.
- *Randomness* represents variations due to the uncertain nature of the stochastic process. The random component of the hydrologic time series can be classified as *autoregressive* or *purely random*.

In statistical modeling of time series, it is assumed that the time series is purely random; therefore, steps should be taken to remove trends, jumps, and periodicity from the data.

5.3 HYDROLOGIC TIME SERIES MODELING: BASIC STEPS

A systematic approach to hydrologic time series modeling could consist of the following main steps (Salas et al., 1988):

- *Data preparation*: The first step is to remove trends, periodicity, and outlying observations and fit the data to a specific distribution by applying the proper transformations.
- *Identification of model composition*: The next step is to decide upon the use of a univariate or multivariate model, or a combination, with disaggregation models. This decision can be made based on the characteristics of the water resources system and existing information.
- *Identification of model type*: Various types of models, such as AR (autoregressive), ARMA (autoregressive moving average), and ARIMA (autoregressive integrated moving average), can be selected in this step; details of these models are presented in this chapter. Statistical characteristics of the time series and the modeler input and knowledge about the types of models are the key factors in identification of model type.
- *Identification of model form*: The form of the selected model should be defined based on the statistical characteristics of the time series such as seasonal and non-seasonal characteristics. The periodicity of the data and how it can be considered in the structure of the selected model is the main issue in this step.
- *Estimation of model parameters*: Several methods, such as method of moments and method of maximum likelihood, can be used for estimating the model parameters.
- *Testing the goodness of fit of the model*: In this step, assumptions such as test of independence and normality of residuals should be checked. The statistics used for verifying these assumptions are briefly explained in this chapter.
- *Evaluation of uncertainties*: When evaluating uncertainties, the model, parameter, and natural uncertainties in data should be analyzed separately. Model uncertainty may be evaluated by testing whether there are significant differences between the statistics of actual data and forecasted and generated data by alternative models. Parameter uncertainty may be determined from the distribution of estimated parameters.

5.4 DATA PREPARATION

The main tasks in data preparation phase can be summarized as:

1. Trends removal
2. Removal of outlying observations
3. Removal of periodicity
4. Fit a well-known distribution to the data by applying proper transformations if needed.

The principal assumption in using statistical models for hydrologic time series is that the variable is considered to be absolutely random. Therefore, if the series contains unnatural components, such as regulated release from reservoirs or trends, they should be removed first. Trend and seasonality are usually detected by the time sequence plot of the series. They can also be characterized by sample autocorrelation functions, which are slowly decaying and nearly periodic.

For example, consider a nonseasonal series, x_t, which contains trend components as follows:

$$x_t = T_t + z_t \tag{5.1}$$

where T_t and z_t are trend and random components, respectively. The trend component can usually be defined as a function of time as:

$$T_t = f\left(t, \cdots, t^m, \beta_1, \cdots, \beta_m\right) \tag{5.2}$$

where (β_1, $\cdots$, β_m) are the parameters to be estimated. These parameters can be estimated based on the least-square method by minimizing the following sum:

$$\sum_t \left(x_t - T_t\right)^2 \tag{5.3}$$

For this purpose, the partial derivatives of the above sum with respect to each of the parameters, (β_1, $\cdots$, β_m), is set to zero. The random series can then be obtained by subtracting the trend from the observed data using Eq. (5.1).

Example 5.1

A series of annual water consumption in a city for a period of 10 years is shown in Table 5.1. Estimate a linear trend function and generate the random series of water consumption by removing the trend.

Solution: The linear trend function can be defined as:

$$T_t = \alpha\, t + \beta \tag{5.4}$$

In the least-square method, the following sum should be minimized:

$$\sum_{t=1}^{10} \left(x_t - T_t\right)^2 = \sum_{t=1}^{10} \left(x_t - \alpha\, t - \beta\right)^2$$

TABLE 5.1
Annual Water Consumption in the City for Example 5.1

Year	Water Consumption (million m^3/year)
1	53
2	62
3	63
4	73
5	79
6	91
7	95
8	105
9	122
10	131

α and β can be calculated as:

$$\alpha = \frac{10\sum_{t=1}^{10} tx_t - \sum_{t=1}^{10} t \sum_{t=1}^{10} x_t}{10\sum_{t=1}^{10} t^2 - \left(\sum_{t=1}^{10} t\right)^2} = 8.55$$

$$\beta = \bar{x} - \alpha\bar{t} = 40.4$$

The random component after removing trend is shown in Table 5.2.

TABLE 5.2
Random Components of Annual Water Consumption for Example 5.1

Year	Water Consumption (million m^3/year)	Trend Component	Random Component
1	53	48.95	53 – 48.95=4.05
2	62	57.5	4.5
3	63	66.05	–3.05
4	73	74.60	–1.6
5	79	83.15	–4.15
6	91	91.70	–0.7
7	95	100.25	–5.25
8	105	108.80	–3.8
9	122	117.35	4.65
10	131	125.90	5.1

The method used in Example 5.1 is known as *classical decomposition* of the series into a trend component, a seasonal component, and a random component. Another method for removing trend and seasonality is *differencing*, which is widely used in modeling of hydrologic time series. For this purpose, the first differencing operator, ∇, is defined as:

$$\nabla(x_t) = x_t - x_{t-1} = (1 - B)x_t \tag{5.5}$$

where B is the backward shift operator,

$$B(x_t) = x_{t-1} \tag{5.6}$$

Higher orders of the operators B and ∇ are defined as:

$$B^j(x_t) = x_{t-j} \qquad \text{and} \qquad \nabla^j(x_t) = \nabla\left(\nabla^{j-1}(x_t)\right), \quad j \geq 1 \tag{5.7}$$

If the first-order difference operator is applied to the linear trend function shown in Eq. (5.4), then we obtain the constant function $\nabla(T_t) = \alpha$. In a similar way, any polynomial trend of degree k can be reduced to a constant by application of the operator ∇^k. Therefore, if a polynomial trend is detected in a hydrologic time series as:

$$T_t = \sum_{i=0}^{m} a_i \, t^i \tag{5.8}$$

Then, by successive differencing of Equation 5.8, the following expression can be obtained:

$$\nabla^m(x_t) = m!\, a_m + \nabla^m z_t \tag{5.9}$$

where z_t is the random component.

In most of the parameter estimation methods, it is assumed that the time series probability distribution is normal, but in many cases the hydrologic time series do not follow normal distribution, are asymmetrically distributed, or are bounded by zero. Therefore, it is often necessary to transform those variables to normal before performing statistical modeling. Different transformations have been used in hydrologic time series analysis. For example, it has been found that the log-normal distribution fits the annual series of precipitation and runoff in the United States (Salas et al., 1988). If we represent these series by X, the following transformation yields the normal series Y:

$$Y = \log(X) \tag{5.10}$$

If gamma probability distribution function is used for the annual series X, then the following transformation yields a relatively normal distribution of Y:

$$Y = \sqrt{X} \tag{5.11}$$

The following two-parameter transformation (first developed by Box and Cox, 1964) has been used by many investigators:

$$z_t = \frac{\left(x_t + \lambda_2\right)^{\lambda_1} - 1}{\lambda_1 g^{(\lambda_1 - 1)}} \qquad \lambda_1 > 0 \tag{5.12}$$

and

$$z_t = g \cdot \ln\left(x_t + \lambda_2\right) \qquad \lambda_1 = 0 \tag{5.13}$$

where λ_1 and λ_2 are the transformation parameters, g is the geometric mean of $x_t + \lambda_2$, and x_t and z_t are the original and transformed series, respectively. The first parameter, λ_1, governs the strength of the transformation. $\lambda_1 = 1$ corresponds to the original data and $\lambda_1 = 0$ to a logarithm. Values are scaled by the geometric mean to keep the variance constant, thus allowing for a first comparison of mean-squared errors between two different transformations.

Hashino and Delleur (1981) developed the following transformation similar to that of Box and Cox:

$$z_t^{(\lambda)} = \frac{\left(x_t + c\right)^{\lambda} - 1}{\lambda} \qquad \lambda \neq 0 \tag{5.14}$$

and

$$z_t^{(\lambda)} = \ln\left(x_t + c\right) \qquad \lambda = 0 \tag{5.15}$$

where λ and c correspond to λ_1 and λ_2 in the Box–Cox transformation. The values in the transformation are not scaled by the geometric mean and therefore do not permit direct comparison of mean-squared errors between two different transformations. Delleur and Karamouz (1982) applied the following transformation on annual data for the Gunpowder and Patasco Rivers in Maryland:

$$z_t = \left(x_t + a\right)^b \qquad b \neq 0 \tag{5.16}$$

and

$$z_t = \log(x_t + a) \qquad b = 0 \tag{5.17}$$

where a and b are the transformation parameters.

5.5 METHODS OF PARAMETER ESTIMATION

The methods of moments, least squares, and maximum likelihood are the basic methods used for estimating model parameters and are explained in the following sections.

5.5.1 Method of Moments

Consider a sample (x_1, ⋯, x_N). The nth sample moment is defined as follows:

$$M_n = \left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i^n \tag{5.18}$$

If a model is fitted to a sample, then the population moments of the series generated by this model are functions of model parameters; therefore, the moment parameter estimates can be obtained by equating sample moments and population moments.

Example 5.2

Suppose the following model is used for the sample (x_1, ⋯, x_N):

$$x_t = \alpha \cdot z_t + \beta$$

where z_t is an independent random variable with mean zero and variance one; α and β are the model parameters. Estimate the model parameters using the method of moments.

Solution: The first and second sample moments can be estimated as:

$$M_1 = \left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i$$

and

$$M_2 = \left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i^2$$

The first two population moments of the model are:

$$E[X] = \beta$$

and

$$E\left[X^2\right] = \alpha^2 + \beta^2$$

By equating the first population and sample moments, it can be written that:

$$\hat{\beta} = \left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i$$

which is an estimate of the parameter β. By equating the second moments, an estimate of the second parameter can be obtained:

$$\hat{\alpha} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \hat{\beta}^2}$$

5.5.2 Method of Least Squares

Consider a sample (x_1, $\cdots$, x_N). A model is fitted to the sample as follows:

$$\hat{x}_t = f\left(x_{t-1}, x_{t-2}, \cdots, \alpha_1, \alpha_2, \cdots, \alpha_m\right) + \varepsilon_t \tag{5.19}$$

where ε_t is the model residual and (α_1, α_2, $\cdots$, α_m) are the model parameters. The objective of the least-squares method is to minimize the following sum:

$$\sum_{t=1}^{N}\left(x_t - \hat{x}_t\right)^2 = \sum_{t=1}^{N}\left(x_t - f\left(x_{t-1}, x_{t-2}, \cdots, \hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_m\right)\right)^2 \tag{5.20}$$

To minimize the above sum, all partial derivatives with respect to estimated values of the parameters ($\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_m$) should be equal to zero. Therefore, they can be written as:

$$\frac{\partial\sum_{t=1}^{N}\left(x_t - \hat{x}_t\right)^2}{\partial\hat{\alpha}_1} = 0, \quad \cdots, \quad \frac{\partial\sum_{t=1}^{N}\left(x_t - \hat{x}_t\right)^2}{\partial\hat{\alpha}_m} = 0 \tag{5.21}$$

By solving these equations simultaneously, the model parameters $(\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_m)$ can be estimated.

Example 5.3

Consider that the following model is used for the sample $(x_1, \cdots, x_N)$:

$$x_t = \alpha x_{t-1} + \varepsilon_t$$

Find the model parameter, α, using the least-squares method.

Solution: The sum of square of errors is estimated as:

$$\sum \varepsilon_t^2 = \sum \left(x_t - \alpha x_{t-1}\right)^2$$

The partial derivatives of the sum with respect to α can be written as:

$$\frac{\partial \sum \left(x_t - \alpha x_{t-1}\right)^2}{\partial \alpha} = \sum \left(x_{t-1}\left(\alpha x_{t-1} - x_t\right)\right) = 0$$

Then the parameter can be estimated as:

$$\alpha = \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2}$$

5.5.3 Method of Maximum Likelihood

Consider the model shown in Equation 5.2. The joint probability of $(\varepsilon_1, \cdots, \varepsilon_N)$ is the *likelihood function* and it can be shown as:

$$L(\varepsilon) = \prod_{t=1}^{N} f\left(\varepsilon_t, \alpha_1, \cdots, \alpha_m\right) \tag{5.22}$$

The maximum likelihood of the parameters $(\alpha_1, \cdots, \alpha_m)$ is obtained by maximizing the function $L(\varepsilon)$. For this purpose, the logarithm of likelihood function can also be maximized:

$$\mathrm{Log}\left(L(\varepsilon)\right) = \log \prod_{t=1}^{N} f\left(\varepsilon_t, \alpha_1, \cdots, \alpha_m\right) = \sum_{t=1}^{N} \log\left[f\left(\varepsilon_t, \alpha_1, \cdots, \alpha_m\right)\right] \tag{5.23}$$

For maximizing the above sum, partial derivatives with respect to the parameters ($\alpha_1, \cdots, \alpha_m$) should be equated to zero:

$$\frac{\partial \mathrm{Log}\big(L(\varepsilon)\big)}{\partial \alpha_1} = 0, \cdots, \frac{\partial \mathrm{Log}\big(L(\varepsilon)\big)}{\partial \alpha_m} = 0 \tag{5.24}$$

The maximum likelihood estimate of the parameters can be obtained by simultaneously solving the above equations.

5.6 GOODNESS OF FIT TESTS

The goodness of fit tests presented in this section are used to identify the nature of the unknown frequency distribution of the population. This procedure determines if a particular shape for the population frequency curve is consistent with the sample results obtained. Two of the most common test statistics for the goodness of fit are:

1. *Chi-square test*, for testing the hypothesis of goodness of fit of a given distribution
2. *Kolmogorov–Smirnov statistic*, for estimating the maximum absolute difference between the cumulative frequency curve of sample data and the fitted distribution function

The following steps should be taken for hypothesis testing (Lapin, 1990):

1. Formulate the null hypothesis (H_0). The data represent a specific distribution.
2. Select the test procedure and test statistic (such as chi-square or Kolmogrov–Smirnov statistics).
3. Select the significance level and the acceptance and rejection regions for the decision rule.
4. Compute the value of the test statistic for the sample data.
5. Make the decision.

5.6.1 Chi-Square Goodness of Fit Test

In this test, two sets of frequencies are compared. First, the data should be divided into a numbers of categories and then the *actual frequencies* (f_i) in each category (class) should be compared with the *expected frequencies* ($\hat{f}_i$), which are estimated based on the fitted distribution. The test statistic can be estimated as follows:

$$\chi^2 = \sum_{i=1}^{NC} \frac{\left(f_i - \hat{f}_i\right)}{\hat{f}_i} \tag{5.25}$$

TABLE 5.3
Precipitation Data for Example 5.4

Interval (i)	Precipitation Depth Classes (mm)	Actual Frequency (f_i)
1	$0 \le P < 5$	38
2	$5 \le P < 10$	26
3	$10 \le P < 15$	12
4	$15 \le P < 20$	10
5	$20 \le P < 25$	8
6	$25 \le P < 30$	3
7	$30 \le P < 35$	2
8	$35 \le P < 40$	1
9	$40 \le P$	0

where NC is the total number of categories. This estimation is close enough for testing the goodness of fit whenever the expected frequency in any category is equal or greater than 5 (Lapin, 1990). χ^2 is chi-square distributed with $NC - NP - 1$ degree of freedom, where NP is the number of parameters to be estimated.

Example 5.4

The classified monthly precipitation data measured in a station are divided into nine categories (see Table 5.3). Previous studies have shown that in spite of assumptions about distribution of precipitation data, it is exponentially distributed in this station. Therefore, the probability that the monthly rainfall depth, P, is equal or below x can be estimated as:

$$\text{Prob}[P \le x] = 1 - e^{-\lambda x}$$

Assume λ to be 0.105. Use the chi-square test with a 5% significance level and comment on the selected distribution for the sample data.

Solution: The expected frequencies are estimated using the selected exponential distribution as shown in the third column of Table 5.2. For example, for the first class of the data, the probability of having rain with depth less than 5 mm is:

$$\text{Prob}[P < 5] = 1 - e^{-.105 \times 5} = 0.41$$

The expected frequencies are then estimated based on the difference between cumulative probabilities (see Table 5.4). For example, for the second interval, it can be written that:

$$\text{Expected frequency} = (0.65 - 0.41) \times 100 = 24$$

TABLE 5.4
Estimated Expected Frequencies and Chi-Square Statistic for the Data Presented in Example 5.4

Interval (i)	Actual Frequency (f_i)	Exponential Cumulative Probability at Upper Limit	Expected Frequencies $\left(\hat{f}_i\right)$	$f_i - \hat{f}_i$	χ^2
1	38	0.41	41	–3	0.22
2	26	0.65	24	2	0.17
3	12	0.79	14	–2	0.29
4	10	0.88	9	1	0.11
5	8	0.93	5	3	1.8
6	3	0.96	3		
7	2	0.97	1	–1	0.14
8	1	0.99	2		
9	0	1	1		
Sum	100	—	100	—	2.73

As shown in Table 5.4, the actual frequency for the intervals 6 to 9 is less than 5; therefore, these classes are grouped together. The chi-square statistic is then estimated as 2.73. The degree of freedom for the test is estimated as:

$$\text{Degree of freedom} = 6 - 1 - 1 = 4$$

From Appendix A and considering a 5% significance level and 4 degrees of freedom, the critical value is $\chi^2_{0.05(4)} = 9.488$. Because the estimated value of the test statistic is smaller than the critical value, the null hypothesis that the monthly precipitation is exponentially distributed is accepted.

5.6.2 Kolmogorov–Smirnov Goodness of Fit Test

The chi-square test performs better when the sample size is large enough to include at least five observations in each class interval. For smaller sample sizes, the Kolmogorov–Smirnov test performs better. This test is based on the maximum difference between actual and expected frequencies. The test statistic can be formulated as follows:

$$D = \max\left|F(x) - \hat{F}(x)\right| \tag{5.26}$$

where F is the actual cumulative relative frequency and $\hat{F}$ is the expected cumulative relative frequency.

The same steps should be taken for hypothesis testing as discussed in Section 5.5. The estimated value for the test statistic should then be compared with the critical value selected based on the significance level and number of the data points.

TABLE 5.5
Estimated Expected Frequencies and Kolmogorov–Smirnov Statistic for the Data Presented in Example 5.5

Interval (i)	Actual Frequency (f_i)	Actual Cumulative Relative Frequency ($F(x)$)	Expected Cumulative Relative Frequency $\left(\hat{F}(x)\right)$	D
1	38	0.38	0.41	0.03
2	26	0.64	0.65	0.01
3	12	0.76	0.79	0.03
4	10	0.86	0.88	0.02
5	8	0.94	0.93	0.01
6	3	0.97	0.96	0.01
7	2	0.99	0.97	0.02
8	1	1	0.99	0.01
9	0	1	1	0

Example 5.5

Solve Example 5.4 using the Kolmogorov–Smirnov goodness of fit test.

Solution: Table 5.5 shows the actual and expected cumulative relative frequencies.

The last column in this table shows the absolute values of the difference between the actual and the expected cumulative frequencies. As it can be seen, the maximum value for the test statistic, D, is estimated as 0.03. From Appendix B and considering 5% significance level and n =100, the critical value, D, is equal to 0.12. Because the estimated value of the test statistic is smaller than the critical value, the null hypothesis that the monthly precipitation is exponentially distributed is accepted.

5.6.3 Tests of Normality

The *graphical test* is the most commonly used test for testing the hypothesis that a given time series is normal. In this test, the empirical distribution of the series is plotted on normal probability paper to determine whether the plotted points can be approximated as a straight line. Figure 5.1 shows the normal probability plot for streamflows of Salt River in northern Arizona. The streamflow data do not follow normal distribution.

The chi-square test has also been widely used for testing the normality of data. The test statistic and the procedure for testing the goodness of fit hypothesis were presented in Section 5.5.1. Another normality test that is widely used in practice is the *skewness test*. The skewness coefficient of a time series x_t is estimated as follows:

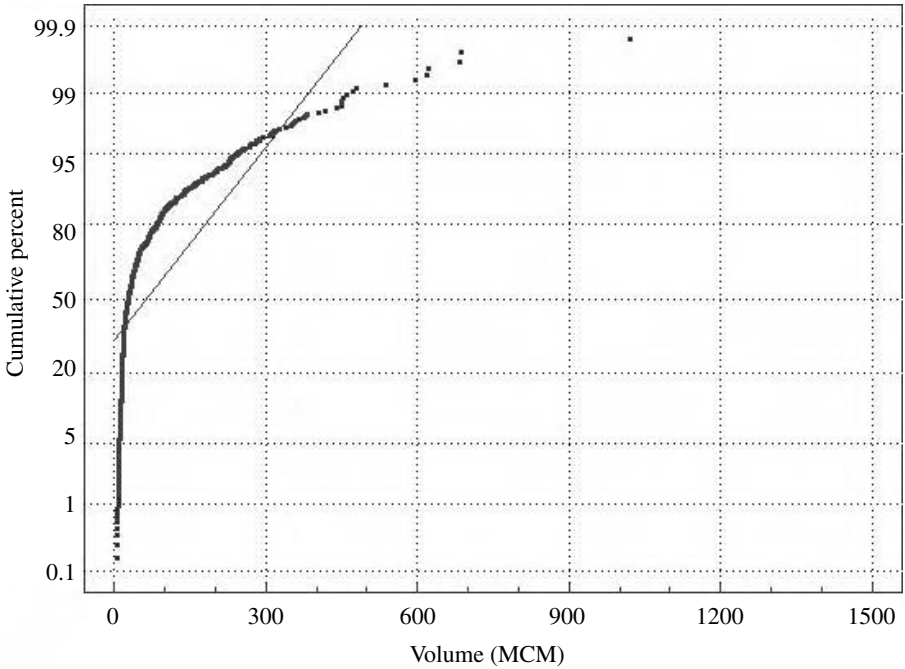


FIGURE 5.1 The normal probability plot for monthly streamflow at Salt River Station on Salt River in Arizona (1914–1998).

$$\hat{\gamma} = \frac{\frac{1}{N}\sum_{t=1}^{N}\left(x_t - \bar{x}\right)^3}{\left[\frac{1}{N}\sum_{t=1}^{N}\left(x_t - \bar{x}\right)^2\right]^{3/2}} \tag{5.27}$$

where N is the total number of sample data and $\bar{x}$ is the sample mean. The skewness test is based on the fact that the skewness coefficient of a normal variable is zero. If the series is normally distributed, $\hat{\gamma}$ is asymptotically normally distributed with mean zero and variance $6/N$ (Snedecor and Cochran, 1967). Then $(1 - \alpha)$ probability limits on γ could be defined as:

$$\left[-u_{1-\alpha/2}\sqrt{6/N} \quad , \quad u_{1-\alpha/2}\sqrt{6/N}\right] \tag{5.28}$$

where $u_{1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution. Therefore, if $\hat{\gamma}$ falls within the limits of Eq. (5.4), the hypothesis of normality is accepted; otherwise, it is rejected. Snedecor and Cochran (1967) recommended the above

TABLE 5.6
Table of Skewness Test for Normality for Sample Sizes, Less than 150

N	α = 0.02	α = 0.01
25	1.061	0.711
30	0.986	0.662
35	0.923	0.621
40	0.870	0.587
45	0.825	0.558
50	0.787	0.534
60	0.723	0.492
70	0.673	0.459
80	0.631	0.432
90	0.596	0.409
100	0.567	0.389
125	0.508	0.350
150	0.464	0.321
175	0.430	0.298

Source: Adapted from Snedecor and Cochran (1967).

procedure when the sample size is greater than 150. For smaller sample sizes, they suggest comparing the computed coefficient of skewness with the tabulated values, $\gamma_\alpha(N)$, presented in Table 5.6. If $|\hat{\gamma}| < \gamma_\alpha(N)$, the hypothesis of normality is accepted.

Example 5.6

In order to bring the probability distribution of the monthly Salt River streamflows (see Figure 5.1) closer to a normal distribution, the logarithmic transformation is applied as follows:

$$y_t = \log_{10}(x_t)$$

where x_t and y_t are the original and transformed streamflow series in month t, respectively. Figure 5.2 shows the normal probability plot for the transformed series for the period 1914 to 1998. The skewness coefficient of the transformed series is estimated to be 0.8. Use the skewness test and comment on the normality of the transformed data.

Solution: As shown in Figure 5.2, the transformed series displays a better fit to the normal distribution compared with the original series; however, it still cannot be approximated as a straight line. If the 5% significance level is selected, the 95% probability limits on the skewness coefficient can be estimated using Eq. (5.4):

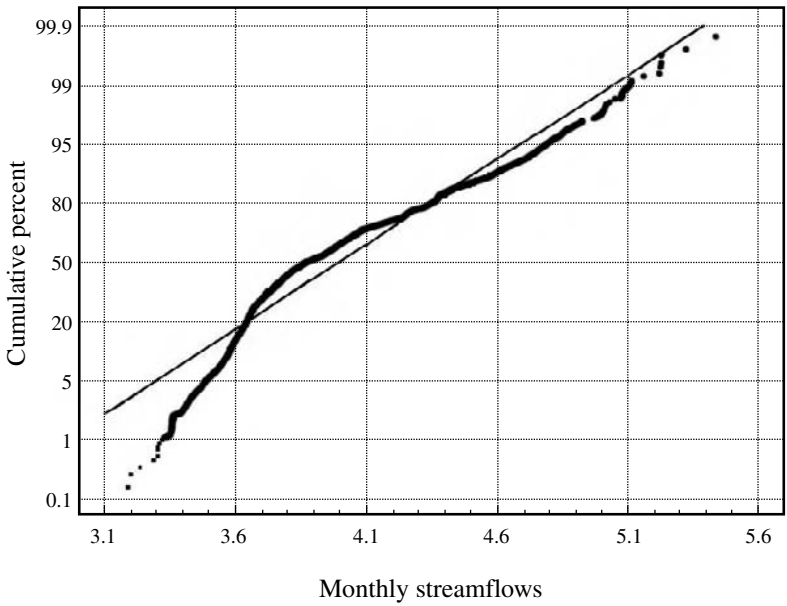


FIGURE 5.2 The normal probability plot for transformed montly streamflow at Salt River Station at Salt River in Arizona (1914–1998).

$$\left[-u_{0.975}\sqrt{6/N} \quad , \quad u_{0.975}\sqrt{6/N}\right] = \left[-1.96\sqrt{6/1020} \quad , \quad 1.96\times\sqrt{6/1020}\right] = \left[-0.15\,,\,0.15\right]$$

Because 0.8 does not fall within the above limits, the series is not normally distributed.

5.6.4 Tests of Independence in Time

The modeling of a hydrologic time series usually assumes that the stochastic component, after removing the periodic components and time-dependence structure are removed, is an independent and normally distributed series (Salas et al., 1988). The basic statistical representation of the time-dependence structure is the correlogram, which shows the fluctuations of the autocorrelation coefficient of the series for different lags. The autocorrelation coefficient is a dimensionless measure of linear dependence that can be estimated as follows:

$$r_k = \frac{\sum_{t=1}^{N-k}\left(x_t - \bar{x}\right)\left(x_{t+k} - \bar{x}\right)}{\sum_{t=1}^{N}\left(x_t - \bar{x}\right)^2} \tag{5.29}$$

where r_k is the lag-k autocorrelation coefficient and x is the sample mean. Figure 5.3 shows the correlogram of the Salt River monthly stream flows for the period of 1914

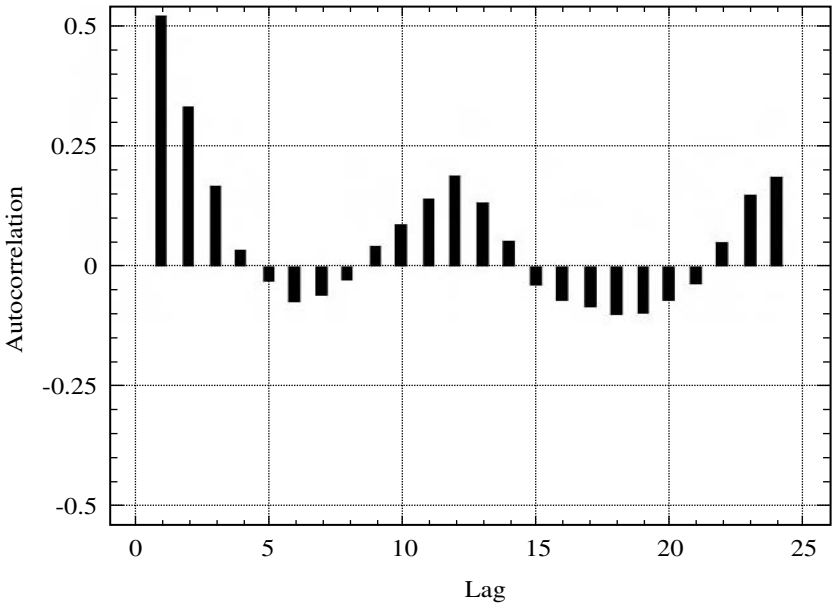


FIGURE 5.3 Correlogram of monthly streamflow of Salt River in Arizona (1914–1998).

to 1998. For an independent series, the population correlogram is equal to zero for $k \neq 0$. Even though correlograms of samples of independent series show fluctuations around zero, they are not necessarily equal to zero due to sampling variability. Anderson (1942) determined probability limits for the correlogram of independent series as follows:

- **Anderson test of independence in time.** Anderson (1942) gave the following limits with 95% and 99% probability levels:

$$r_k(95\%) = \frac{-1 \pm 1.645\sqrt{N-k-1}}{N-k} \tag{5.30}$$

$$r_k(99\%) = \frac{-1 \pm 2.326\sqrt{N-k-1}}{N-k} \tag{5.31}$$

  where N is the sample size. Based on this test, autocorrelations falling outside of the probability limit indicate significant time dependence; otherwise, the series can be considered as an independent series.
- **Porte Manteau test.** The Porte Manteau lack-of-fit test was first developed by Box and Pierce (1970) as an approximate test of model adequacy. Hipel and McLeod (1977) and others applied this test for verifying linear

models of hydrologic time series. In this test, instead of checking to see if autocorrelations fall within the confidence bound, a single statistic that depends on the autocorrelation function is used. Suppose that the autoregressive integrated moving average (ARIMA) (p, d, q) model (discussed later in this chapter) is selected for representing a time series, where p is the number of autoregressive terms, q is the number of moving average terms, and d is the number of differences. As mentioned, this test evaluates the adequacy of a model by determining whether the residual series of the model is independent. The following statistic is used in this test:

$$Q = (N - d) \cdot \sum_{k=1}^{L} r_k^2(\varepsilon) \tag{5.32}$$

where L is maximum lag considered. The statistic Q is approximately chi-square distributed with $L - p - q$ degrees of freedom. Therefore, if the estimated test statistic is smaller than the chi-square value, $\chi_\alpha^2(L - p - q)$, of a given probability limit $(1 - \alpha)$, ε_t is an independent series and the model is adequate. Ljung and Box (1978) argued that under the hypothesis of model adequacy the cutoff value given by $\chi_\alpha^2\ (L - p - q)$ is closer to the $(1 - \alpha)$ quantile of the distribution of the following statistic:

$$Q' = n(n+2)\sum_{k=1}^{L} r_k^2(\varepsilon)/(n-j) \tag{5.33}$$

Tao and Delluer (1976) used the following statistic for testing the adequacy of a seasonal ARMA model:

$$Q_1 = N \cdot \sum_{k=1}^{L}\sum_{\tau=1}^{\omega} r_{k,\tau}^2(\varepsilon) \tag{5.34}$$

where ω is the number of time intervals per year (seasons). Granger and Anderson (1978) found examples where the residuals were uncorrelated while the squared residuals were correlated. The sample autocorrelation function of squared residuals $\hat{\varepsilon}_t^2$ can be estimated as (Brockwell and Davis, 1987):

$$\hat{\rho}_{\varepsilon\varepsilon}(L) = \frac{\sum_{t=1}^{N-L}\left(\hat{\varepsilon}_t^{\,2} - \overline{\varepsilon^2}\right)\left(\hat{\varepsilon}_{t+L}^2 - \overline{\varepsilon^2}\right)}{\sum_{t=1}^{N}\left(\hat{\varepsilon}_t^2 - \overline{\varepsilon^2}\right)} \qquad L \geq 1 \tag{5.35}$$

where

$$\overline{\varepsilon^2} = \frac{\sum_{t=1}^{N} \hat{\varepsilon}_t^2}{N}$$

McLeod and Li (1983) showed that:

$$Q_{\varepsilon\varepsilon} = N(N+2)\sum_{k=1}^{L} \hat{\rho}_{\varepsilon\varepsilon}^2(k)/(N-k) \tag{5.36}$$

has an approximate $\chi^2(L)$ distribution under the assumption of model adequacy; therefore, the adequacy of model is rejected at significance level α if:

$$Q_{\varepsilon\varepsilon} > \chi_{1-\alpha}^2(L) \tag{5.37}$$

- **Cumulative periodogram test.** This test can be used for determining whether the periodicity is adequately removed from series. The periodogram of sample residual, $\hat{\varepsilon}_t$, can be estimated as follows (Salas et al., 1988) and MSD stands for mean squared division:

$$\mathrm{MSD}(h_j) = \frac{2}{N^2}\left[\left(\sum_{t=1}^{N} \hat{\varepsilon}_t \cdot \mathrm{Cos}(2\pi h_j t)\right)^2 + \left(\sum_{t=1}^{N} \hat{\varepsilon}_t \mathrm{Sin}(2\pi h_j t)\right)^2\right] \tag{5.38}$$

where $h_j = j/N$ is the harmonic (frequency). The cumulative periodogram of $\hat{\varepsilon}_t$ can then be defined as:

$$P_i = \frac{\sum_{j=1}^{i} \mathrm{MSD}(h_j)}{\sigma_\varepsilon^2} \qquad i = 1, \cdots, N/2 \tag{5.39}$$

where σ_ε^2 is the variance of $\hat{\varepsilon}_t$. For an independent series, the plot of P_i versus i will be scattered around a line joining (0,0) and (0.5N,1). A periodicity with harmonic h_j in the residuals would produce a large value of MSD(h_j) and in turn would appear as a deviation from the straight line, which is valid for an independent series (Salas et al., 1988). Probability limit lines using the Kolmogorov–Smirnov statistic can be drawn to test the significance of such deviations (Box and Jenkins, 1970). Considering the significance level α, the limit lines for a real independent series are drawn approximately at distances $\pm K_\alpha / \sqrt{N'}$ around the theoretical

TABLE 5.7
Approximate Critical Values of the Kolmogorov–Smirnov Statistic K_α (Yevjevich 1972a)

α	0.01	0.05	0.1	0.2	0.25
K_α	1.63	1.36	1.22	1.07	1.02

straight line where $N' = (N - 2)/2$ or $N' = (N - 1)/2$ for even or odd N values, respectively. If more than αN of the plotted points fall outside the probability lines, the residual series still has some periodicity; otherwise, it can be concluded that the residuals are independent. Approximate values of K_α are given in Table 5.7.

- **Turning point test of randomness.** Suppose the time series x_t are observations of a random sequence. The turning point at time t is detected when $x_{t-1} < x_t$ and $x_t > x_{t+1}$ or $x_t < x_{t-1}$ and $x_t < x_{t+1}$. If T_p is the number of turning points, then the probability of a turning point at time t is 2/3. The expected number of turning points as stated by Brockwell and Davis (1987) is:

$$\mu_{T_p} = \frac{2(n-2)}{3} \tag{5.40}$$

The variance of the number of turning points can be determined as follows:

$$\sigma^2_{T_P} = \frac{(16n-29)}{90} \tag{5.41}$$

It can be shown for a random sequence, the number of turning points (T_p) has an $N(\mu_{T_p}, \sigma^2_{T_p})$ distribution. So the assumption that x_t is a random sequence is rejected if:

$$\frac{\left|T_p - \mu_{T_p}\right|}{\sigma_{T_p}} > u_{1-\alpha/2} \tag{5.42}$$

where $u_{1-\alpha/2}$ is the $(1 - \alpha/2)$ probability level of a standard normal distribution.

Example 5.7

A series of residuals of a model fitted to precipitation data in a station for 20 time intervals is given in Table 5.8. Use the turning point test and comment on the randomness of the residuals.

TABLE 5.8
The Series of Residuals for Example 5.7

Month	Residual	Month	Residual
1	0.87	11	0.30
2	0.45	12	0.51
3	0.33	13	0.41
4	0.36	14	0.24
5	0.31	15	0.56
6	0.12	16	0.48
7	0.33	17	0.06
8	0.17	18	0.77
9	0.78	19	0.36
10	0.90	20	0.22

Solution: The turning points occurred in months 3, 4, 6, 7, 8, 10, 11, 12, 14, 17, 18. The expected value and variance of the number of turning points can be estimated using Eqs. (5.16) and (5.17):

$$\mu_{T_p} = \frac{2(20-2)}{3} = 12 \quad \text{and} \quad \sigma_{T_p} = \sqrt{\frac{(16\times 20-29)}{90}} = 1.80$$

The statistic for testing the randomness can be defined using Eq. (5.18):

$$\frac{|11-12|}{1.8} = 0.556$$

which is less than the critical value of the $N(0,1)$ distribution with a 5% significance level ($u_{1-0.05/2} = 1.96$), so the assumption of randomness is accepted.

5.7 AKAIKE'S INFORMATION CRITERION (AIC)

Akaike's information criterion (AIC) (Akaike, 1974), which considers parsimony in model building, is the primary criterion used for model selection. Parsimony is the tendency to select minimum number of parameters in order to reduce the chance of error accumulation. Akaike defined the AIC among competing autoregressive moving average (ARMA (p,q)) models as follows:

$$\mathrm{AIC}(p,q) = N \cdot \ln\left(\hat{\sigma}^2(\varepsilon)\right) + 2(p+q) \tag{5.43}$$

where N is the sample size and $\hat{\sigma}^2(\varepsilon)$ is the maximum likelihood estimate of the residual variance. Based on this criterion, the model with minimum AIC is selected.

The goodness of fit tests should be applied to the model with minimum AIC to make sure the residuals are consistent with their expected behavior. If the model residuals do not pass the test, the models with higher values of AIC should be checked.

5.8 DATA GENERATION AND FORECASTING

Two major uses of time-series modeling are generation of synthetic values and forecasting future events. The generation of synthetic time series often consists of the generation of independent normal variables in a series defined as a random process. The generated random component of a series will be added to its trend or deterministic component to reproduce the original time series. The second important application of time series models is forecasting of future events. The forecast values are often different from the observed values; therefore, a confidence interval is assigned for the forecast values.

5.9 AUTOREGRESSIVE MODELING

Autoregressive (AR) models have been used widely in hydrologic time-series modeling. These models incorporate the correlation between time sequences of variables. These models are the simplest models and their development goes back to the application of Thomas Fiering and Markov lag-1 models. They can be classified into the following subsets:

- AR models with constant parameters, which are typically used for modeling of annual series
- AR models with timely variable parameters, which are typically used for modeling of seasonal (periodic) series

The basic form of the AR model of order p, (AR p), with constant parameters is:

$$z_t = \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t \tag{5.44}$$

where:

z_t is the time-dependent normal and standardized series $\{N(0,1)\}$. ϕ_i are the autoregressive coefficients. ε_t is the time-independent variable (white noise). p is the order of the autoregressive model.

The above formulation is based on the standardized series, which can be obtained as follows:

$$z_t = \frac{x_t - \mu}{\sigma} \tag{5.45}$$

where μ and σ are the mean and standard deviation of the series x_t. The parameter set of the model is:

$$\left\{\mu, \sigma, \phi_1, \cdots, \phi_p, \sigma^2(\varepsilon)\right\} \tag{5.46}$$

where $\sigma^2(\varepsilon)$ is the variance of the time-independent series. The model parameters can be estimated by solving the following linear equations (Yule–Walker equations) simultaneously:

$$r_i = \hat{\phi}_1 r_{i-1} + \hat{\phi}_2 r_{i-2} + \cdots + \hat{\phi}_p r_{i-p} \quad (i \le p) \tag{5.47}$$

where r_i are the sample correlation coefficients of lag i. The parameter $\sigma^2(\varepsilon)$ can also be estimated using the following relation:

$$\hat{\sigma}^2(\varepsilon) = \frac{N\hat{\sigma}^2}{(N-p)}\left(1 - \sum_{i=1}^{p} \hat{\phi}_i r_i\right) \tag{5.48}$$

where N is the number of the data and $\hat{\sigma}^2$ is the sample variance.

The stationary condition must be met by the model parameters. For this purpose, the roots of the following equation should lie inside the unit circle (Yevjevich, 1972):

$$u^p - \hat{\phi}_1 u^{p-1} - \hat{\phi}_2 u^{p-2} - \cdots - \hat{\phi}_p = 0 \tag{5.49}$$

In other words, we must have $|u_i| < 1$. In order to forecast or generate annual AR models, the following relation can be used:

$$\hat{z}_t = \hat{\phi}_1 \hat{z}_{t-1} + \cdots + \hat{\phi}_P \hat{z}_{t-P} + \hat{\sigma}(\varepsilon)\zeta_t \tag{5.50}$$

where ζ_t is the standardized normal variable.

Example 5.8

For an AR(2) model, the parameters have been estimated as $\phi_1 = 0.5$ and $\phi_2 = 0.3$. Check the stationary condition of the parameters.

Solution: Using Eq. (5.49), the following expressions can be written:

$$u^2 - 0.5u - 0.3 = 0 \quad \Rightarrow \quad \begin{cases} u_1 = 0.85 \\ u_2 = -0.35 \end{cases}$$

The roots lie within the unit circle; therefore, the parameters pass the stationary condition.

Example 5.9

For a sample of 100-year normal and standardized annual inflows to a reservoir, an AR(2) model is selected. The first and second correlation coefficients are estimated as $r_1 = 0.6$ and $r_2 = 0.35$. Estimate the model parameters if the variance of the normal and standardized inflow series is estimated as 1.5.

Solution: Using Eq. (5.47), we write:

$$\begin{cases} r_1 = \phi_1 + \phi_2 r_1 \\ r_2 = \phi_2 + \phi_1 r_1 \end{cases}$$

By solving these equations,

$$\phi_1 = \frac{r_1(1-r_2)}{1-r_1^2} \quad \text{and} \quad \phi_2 = \frac{r_2 - r_1^2}{1-r_1^2}$$

Then,

$$\phi_1 = 0.61 \quad \text{and} \quad \phi_2 = -0.02$$

The variance of model residuals can also be estimated using Eq. (5.48):

$$\hat{\sigma}^2(\varepsilon) = \frac{100 \times 1.5}{(100-2)}(1 - 0.6 \times 0.61 + 0.35 \times 0.02) = 0.96$$

In order to identify whether an AR(p) is an appropriate model for a specific time series, it is necessary to estimate and investigate the behavior of the *partial autocorrelation function* (PACF) of the series. The PACF represents the time-dependence structure of the series in a different way. The partial autocorrelation lag k may be regarded as the correlation between z_1 and z_{k+1}, adjusted for the intervening observations (z_2, …, z_k). For an AR(p) process, the partial autocorrelation function is the last autoregressive coefficient of the model $\phi_k(k)$, which can be estimated by rewriting Eq. (5.33) as follows:

$$r_i = \hat{\phi}_1(k)\, r_{i-1} + \hat{\phi}_2(k)\, r_{i-2} + \cdots + \hat{\phi}_k(k)\, r_{i-k} \quad (i = 1, \ldots, k)$$

which represents a set of linear equations,

$$\begin{aligned} \hat{\phi}_1(k)\, r_0 + \hat{\phi}_2(k)\, r_1 + \cdots + \hat{\phi}_k(k)\, r_{k-1} &= r_1 \\ \hat{\phi}_1(k)\, r_1 + \hat{\phi}_2(k)\, r_0 + \cdots + \hat{\phi}_k(k)\, r_{k-1} &= r_2 \\ &\vdots \\ \hat{\phi}_1(k)\, r_{k-1} + \hat{\phi}_2(k)\, r_{k-2} + \cdots + \hat{\phi}_k(k)\, r_0 &= r_k \end{aligned}$$

which can be written as:

$$\begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{k-1} \\ r_1 & 1 & r_1 & \cdots & r_{k-2} \\ r_2 & r_1 & 1 & \cdots & r_{k-3} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varphi_1(k) \\ \varphi_2(k) \\ \varphi_3(k) \\ \vdots \\ \varphi_k(k) \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_k \end{bmatrix}$$

Thus, the partial autocorrelation function $\phi_k(k)$ can be determined by successively applying the above relation. It can be shown that the partial correlogram of an autoregressive process of order p has peaks at lags 1 through p and then cuts off. Hence, the partial autocorrelation function can be used to identify the order p of an AR(p) model.

The basic form of the AR model with periodic parameters is

$$z_{\nu,\tau} = \sum_{i=1}^{P} \phi_{i,\tau} z_{\nu,\tau-i} + \sigma_\tau(\varepsilon)\zeta_{\nu,\tau} \tag{5.51}$$

where $z_{\nu,\tau}$ is normal and the standardized value in year ν and season τ, $\phi_{i,\tau}$ are periodic autoregressive coefficients, and $\sigma_\tau(\varepsilon)$ is the standard deviation of residuals in season τ. $z_{\nu,\tau}$ is estimated using the seasonal mean and variances as follows:

$$z_{\nu,\tau} = \frac{x_{\nu,\tau} - \mu_\tau}{\sigma_\tau} \tag{5.52}$$

where μ_τ and σ_τ are the mean and standard deviation of x in season τ, respectively. The parameter set of the model can be summarized as:

$$\left\{\mu_\tau, \sigma_\tau, \phi_{1,\tau}, \cdots, \phi_{P,\tau}, \sigma_\tau^2(\varepsilon), \tau = 1, \cdots, \eta\right\}$$

where η is the total number of seasons. Equation (5.47) can be similarly used for each season:

$$r_k = \sum_{i=1}^{P} \hat{\phi}_{i,\tau} r_{|k-i|,\tau-\min(k,i)} \quad (k > 0) \tag{5.53}$$

The residual variance can be estimated using the following relation:

$$\hat{\sigma}_\tau^2(\varepsilon) = 1 - \sum_{j=1}^{P} \hat{\phi}_{j,\tau} \cdot \hat{r}_{j,\tau} \tag{5.54}$$

5.10 MOVING AVERAGE PROCESS

The autoregressive models can be used as an effective tool for modeling hydrologic time series such as low-flow-season streamflow that is mainly supplied from groundwater and has low variations. But the previous studies have shown that the streamflows in high-flow season can be better formulated by adding a moving average component to the autoregressive component (Salas et al., 1988).

If the series z_t is dependent only on a finite number of previous values of a random variable, ε_t, then the process can be called a *moving average process*. The moving average model of order q (MA(q)) can be formulated as:

$$z_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \theta_3\varepsilon_{t-3} - \cdots - \theta_q\varepsilon_{t-q} \tag{5.55}$$

It can also be written as:

$$z_t = -\sum_{j=0}^{q} \theta_j \varepsilon_{t-j} \qquad (\theta_0 = -1) \tag{5.56}$$

where $\theta_1, \cdots, \theta_q$ are q orders of MA(q) model parameters. The parameter set of the model can be summarized as:

$$\left\{\mu, \theta_1, \cdots, \theta_q, \sigma^2(\varepsilon)\right\}$$

The following relation can be used for estimating the moving average parameters:

$$r_k = \begin{cases} \dfrac{\sum_{j=0}^{q-k} \theta_j \theta_{j+k}}{1 + \sum_{j=1}^{q} \theta_j^2} & (k \le q) \\ 0 & (k > q) \end{cases} \tag{5.57}$$

The parameters of the model should satisfy the *invertibility* condition. This condition, for example, for a MA(1) process, shows that the first order moving average process is inverted into an autoregressive process of infinite order, provided that $|\theta_1 < 1|$. In other words, the root of $u - \theta_1 = 0$ should be inside the unit circle. To generalize the invertibility condition to an MA(q), the roots of the following polynomial should lie inside the unit circle:

$$u^q - \hat{\theta}_1 u^{q-1} - \hat{\theta}_2 u^{q-2} - \cdots - \hat{\theta}_q = 0 \tag{5.58}$$

Example 5.10

For an MA(2) model, the parameters have been estimated as $\theta_1 = 0.65$ and $\theta_2 = 0.3$. Determine whether the parameters pass the invertibility condition.

Solution: Using Eq. (5.35), we can write:

$$u^2 - 0.65u - 0.3 = 0 \Rightarrow \begin{cases} u_1 = 0.96 \\ u_2 = -0.31 \end{cases} \tag{5.59}$$

The roots lie within the unit circle; therefore, the parameters pass the invertibility condition.

5.11 AUTOREGRESSIVE MOVING AVERAGE MODELING

The ARMA model of order (p,q) can be defined by combining an autoregressive model of order p and a moving average model of order q as follows:

$$z_t - \phi_1 z_{t-1} - \cdots - \phi_p z_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{5.60}$$

The ARMA(p,q) model can also be shown in the following compact form:

$$\phi(B) z_t = \theta(B) \varepsilon_t \tag{5.61}$$

where $\phi(z)$ and $\theta(z)$ are the pth and qth degree polynomials:

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \tag{5.62}$$

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \tag{5.63}$$

The parameter set of the ARMA(p,q) model can be summarized as:

$$\left\{\mu, \theta_1, \cdots, \theta_q, \phi_1, \cdots, \phi_p, \sigma^2(\varepsilon)\right\} \tag{5.64}$$

The parameters of the ARMA(p,q) model should satisfy both the conditions of invertibility and stationarity.

Table 5.9 shows the characteristic behavior of the autocorrelation and of the partial autocorrelation for the AR, MA, and ARMA processes, which can be effectively incorporated in selecting the type and form of the statistical model.

TABLE 5.9
Characteristic Behavior of AR, MA, and ARMA Processes

Process	Autocorrelation	Partial Autocorrelation
AR(p)	Damped and infinite in extent exponentials and/or waves	Peaks at lags 1 through p and then cuts off
MA(q)	Peak at lags 1 through q and then cuts off	Damped and infinite in extent exponentials and/or waves
ARMA(p,q)	Irregular in first q-p lags and then damped and infinite in extent exponentials and/or waves	Irregular in first p-q lags and then damped and infinite in extent exponentials and/or waves

5.11.1 Data Generation and Forecasting Using ARMA Models

Once the ARMA model is fitted to a time series, the following procedures can be used for generating or forecasting the values of that time series. For generation using the ARMA(p,q) model (Eq. (5.60)), it is necessary to give p initial Z values to the model. Equation (5.60) may be used recursively to generate the desired numbers of Z_t. By generating a sufficiently long series and deleting 50 or 100 initial terms, the transient effect of the initial values is negligible (Salas et al., 1988). The synthetic generated values conserve the statistical properties of the historical data.

Also, ARMA(p,q) can be used to forecast Z values for lead time L. If $Z_t(L)$ denotes the value of Z_t at lead time L, the following equations could be used for forecasting:

$$Z_t(L) = \phi_1 Z_{t+L-1} + \phi_2 Z_{t+L-2} + ... + \phi_p Z_{t+L-p} - \theta_1 \varepsilon_{t+L-1} - ... - \theta_q \varepsilon_{t+L-q} \quad \text{for} \quad L \le q \quad (5.65)$$

$$Z_t(L) = \phi_1 Z_{t+L-1} + \phi_2 Z_{t+L-2} + ... + \phi_p Z_{t+L-p} \quad \text{for} \quad L > q \quad (5.66)$$

5.12 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELING

The ARMA models are suitable for the data that have the following two basic characteristics:

1. No apparent deviation from stationarity
2. Rapidly decreasing autocorrelation function

If these conditions are not met by a time series, a proper transformation should be performed to generate time series satisfying the above two conditions. This has usually been achieved by differencing, satisfying the essence of ARIMA models. This class of models is very powerful for describing stationary and nonstationary time series. The nonseasonal form of ARIMA models of order (p,d,q) can be formulated as:

$$\phi(B)(1-B)^d z_t = \theta(B)\varepsilon_t \tag{5.67}$$

where $\phi(B)$ and $\theta(B)$ are polynomials of degree p and q, respectively:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - .... - \phi_p B^p \tag{5.68}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - .... - \theta_q B^q \tag{5.69}$$

The model has $p + q + 1$ parameters to be estimated:

$$\left\{\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma^2(\varepsilon)\right\}$$

In order to model the seasonal hydrologic time series, the seasonal form of ARIMA models of nonseasonal order (p,d,q) and of seasonal order $(P,D,Q)_w$ with seasonality w has also been developed. The general multiplicative ARIMA$(p,d,q)(P,D,Q)_w$ can be formulated as:

$$\begin{aligned}&(1 - \Phi_1 B^w - \Phi_2 B^{2w} - .... - \Phi_p B^{Pw})(1 - \phi_1 B - \phi_2 B^2 - .... - \phi_p B^p)(1 - B^w)^D (1 - B)^d z_t = \\ &(1 - \Theta_1 B^w - \Theta_2 B^{2w} - ... - \Theta_Q B^{QW})(1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)\varepsilon_t\end{aligned} \tag{5.70}$$

where:

ε_t is an independently distributed random variable.
B is the backward operator; $B(z_t) = z_{t-1}$.
$(1 - B^w)^D$ is the Dth seasonal difference of season w.
$(1 - B)^d$ is the dth nonseasonal difference.
p is the order of the nonseasonal autoregressive component.
q is the order of the nonseasonal moving average component.
P is the order of the seasonal autoregressive component.
Q is the order of the seasonal moving average component.
Φ is the seasonal autoregressive parameter.
Θ is the seasonal moving average parameter.
ϕ is the nonseasonal autoregressive parameter.
θ is the nonseasonal moving average parameter.

Equation (5.70) can also be shown in the following condensed form:

$$\Phi\left(B^w\right)\phi(B)\left(1 - B^w\right)^D (1 - B)^d z_t = \Theta\left(B^w\right)\theta(B)\varepsilon_t \tag{5.71}$$

Estimating the parameters are done by computer packages. For details of parameter estimation refer to Salas et al. (1989).

Example 5.11

Formulate the ARIMA$(1,0,1)(1,0,1)_{12}$.

Solution: Using Eq. (5.67), it can be written that:

$$(1-\Phi_1 B^w)(1-\phi_1 B)z_t = (1-\Theta_1 B^w)(1-\theta_1 B)\varepsilon_t$$

or

$$z_t - \phi_1 z_{t-1} - \Phi_1 z_{t-12} + \phi_1\Phi_1 z_{t-13} = \varepsilon_t - \theta_1\varepsilon_{t-1} - \Theta_1\varepsilon_{t-12} - \theta_1\Theta_1\varepsilon_{t-13}$$

The model has nonlinear parameters because of the terms $\theta\Theta$ and $\phi\Phi$.

5.12.1 Time Series Forecasting Using ARIMA Models

The ARIMA models are nonstationary and cannot be used for synthetic generation of stationary time series but they are useful for forecasting. The previous ARIMA model equations can be used for data forecasting. As an example, the forecasting equations for ARIMA$(1,0,1)(1,0,1)_{12}$ could be represented as follows:

$$\begin{aligned} z_t(L) = {} & \phi_1 z_{t-1+L} + \Phi_1 z_{t-12+L} - \phi_1\Phi_1 z_{t-13+L} - \theta_1\varepsilon_{t-1+L} \\ & - \Theta_1\varepsilon_{t-12+L} - \theta_1\Theta_1\varepsilon_{t-13+L} \quad \text{for } L \le 12 \end{aligned} \tag{5.72}$$

$$z_t(L) = \phi_1 z_{t-1+L} + \Phi_1 z_{t-12+L} - \phi_1\Phi_1 z_{t-13+L} \quad \text{for } L > 12 \tag{5.73}$$

5.13 MULTIVARIATE AND DISAGGREGATION MODELING OF TIME SERIES

5.13.1 Multivariate Modeling

The planning, design, and operation of water resources systems often involve several hydrologic time series. In this case, multivariate stochastic analysis and multivariate modeling are necessary to consider cross-correlations between different time series. Many investigators, such as Fiering (1964), Matalas (1967), Mejia (1971), Valencia and Schaake (1973), and O'Connell (1974), have proposed multivariate models.

Disaggregation modeling is often done within a multivariate model framework. Disaggregation models are used to decompose time series to several subseries that are temporal or spatial fractions of the key time series. Most applications of disaggregation have been in the temporal domain, although some investigators have applied the same principle in the spatial domain. The first well-known disaggregation

model was presented by Valencia and Schaake (1973). Multivariate and disaggregation models may be used either for the generation of synthetic time series or for forecasting. For disaggregation models, the key series must be available as an input to the model.

An AR(1) multivariate model may be represented as:

$$Z_t = A_1 Z_{t-1} + B\varepsilon_t \tag{5.74}$$

where Z is an $(n \times 1)$ vector of elements z, which represent the values of n time series; A_1 and B are an $(n \times n)$ matrix of parameters obtained by calculating cross correlation presented in the next section; and ε is an $(n \times 1)$ vector of independent, normally distributed random variables with zero mean and variance one. Equation (5.74) may be written in the matrix form as:

$$\begin{bmatrix} Z_t^{(1)} \\ Z_t^{(2)} \\ \vdots \\ Z_t^{(n)} \end{bmatrix} = \begin{bmatrix} a^{11} & a^{12} & \cdots & a^{n1} \\ a^{21} & a^{22} & \cdots & a^{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a^{n1} & a^{n2} & \cdots & a^{nn} \end{bmatrix} \begin{bmatrix} Z_{t-1}^{(1)} \\ Z_{t-1}^{(2)} \\ \vdots \\ Z_{t-1}^{(n)} \end{bmatrix} + \begin{bmatrix} b^{11} & b^{12} & \cdots & b^{n1} \\ b^{21} & b^{22} & \cdots & b^{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b^{n1} & b^{n2} & \cdots & b^{nn} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \vdots \\ \varepsilon_t^{(n)} \end{bmatrix} \tag{5.75}$$

The correlation structure of Z_t of the above equation indicates a lag-0 and lag-1 cross-correlation in time. In the case of two time series ($n = 2$), the formulations for AR(1) multivariate model are:

$$\left.\begin{aligned} Z_t^{(1)} &= a_{11} Z_{t-1}^{(1)} + a_{12} Z_{t-1}^{(2)} + b_{11}\varepsilon_t^{(1)} + b_{12}\varepsilon_t^{(2)} \\ Z_t^{(2)} &= a_{21} Z_{t-1}^{(1)} + a_{22} Z_{t-1}^{(2)} + b_{21}\varepsilon_t^{(1)} + b_{22}\varepsilon_t^{(2)} \end{aligned}\right\} \tag{5.76}$$

The extended AR(2) multivariate model is:

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + B\varepsilon_t \tag{5.77}$$

The parameters of Eq. (5.72) are matrices A_1, A_2, and B; Salas et al. (1988) provide details of the mathematical estimation of the model parameters.

The main parameter for estimating the other parameters of multivariate models is the covariance structure represented by a correlation matrix. The correlation matrix of lag k for n correlated time series can be obtained as:

$$M_k = \begin{bmatrix} \rho_k^{11}, & \rho_k^{12}, & ..., & \rho_k^{1n} \\ \rho_k^{21}, & \rho_k^{22}, & ..., & \rho_k^{2n} \\ . & . & . & . \\ \rho_k^{n1}, & \rho_k^{n2}, & ..., & \rho_k^{nn} \end{bmatrix} \tag{5.78}$$

Where ρ_k^{ij} is the lag-k cross-correlation coefficient between time series Z_t^i and Z_t^j. For the multivariate AR(1) model, the multivariate correlation function can be written by one of the following two expressions:

$$\left.\begin{array}{l} M_k = A_1 M_{k-1},\ k > 0 \\ \\ \text{or} \\ \\ M_k = A_1^k M_0,\ k \geq 0 \end{array}\right\} \tag{5.79}$$

And, for the multivariate AR(2) model, the multivariate correlation function is:

$$M_k = A_1 M_{k-1} + A_2 M_{k-2},\ k > 0 \tag{5.80}$$

The moment estimates of the parameters A_1 and B for the multivariate AR(1) may be obtained from Eqs. (5.81) and (5.82):

$$\hat{A}_1 = \hat{M}_1 \hat{M}_0^{-1} \tag{5.81}$$

and

$$\hat{B}\ B^T = \hat{M}_0 - \hat{A}_1 \hat{M}_1^T \tag{5.82}$$

Similarly, for the multivariate AR(2) model, the moment estimate of the parameters A_1, A_2, and B may be obtained from Eqs. (5.78) and (5.79) (Salas and Pegram, 1978):

$$\hat{A}_1 = \left[\hat{M}_1 - \hat{M}_2 \hat{M}_0^{-1} \hat{M}_1^T\right] \times \left[\hat{M}_0 - \hat{M}_1 \hat{M}_0^{-1} \hat{M}_1^T\right]^{-1} \tag{5.83}$$

$$\hat{A}_2 = \left[\hat{M}_2 - \hat{M}_1 \hat{M}_0^{-1} \hat{M}_1^T\right] \times \left[\hat{M}_0 - \hat{M}_1^T \hat{M}_0^{-1} \hat{M}_1\right]^{-1} \tag{5.84}$$

$$BB^T = \hat{M}_2 - \left[\hat{A}_1 \hat{M}_1^T + \hat{A}_2 \hat{M}_2^T\right] \tag{5.85}$$

Example 5.12

Table 5.10 provides 18 years of annual flows at two stations of a river basin. Formulate the multivariate model AR(1) for generating synthetic data for these two stations. Assume that the data follow normal distribution.

TABLE 5.10
Annual Streamflows of Two Stations (million m³)

Year	Streamflows of Station 1	Streamflows of Station 2
1	150	250
2	200	450
3	330	502
4	401	720
5	550	550
6	662	600
7	445	465
8	551	650
9	302	390
10	203	300
11	405	520
12	470	740
13	502	780
14	608	705
15	402	600
16	501	590
17	460	480
18	230	320

Solution: The correlation coefficients matrices of lag-0 and lag-1 are computed using Eq. (5.78):

$$M_0 = \begin{bmatrix} 1 & 0.76 \\ 0.76 & 1 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 0.41 & 0.12 \\ 0.52 & 0.37 \end{bmatrix}$$

The parameters A_1 and BB^T are estimated using Eqs. (5.81) and (5.82), respectively:

$$A_1 = \begin{bmatrix} 0.78 & -0.47 \\ 0.56 & -0.05 \end{bmatrix}$$

$$BB^T = \begin{bmatrix} 0.74 & 0.54 \\ 0.54 & 0.73 \end{bmatrix}$$

Matrix B could then be estimated as:

$$B = \begin{bmatrix} 0.82 & 0 \\ 0.6 & -0.57 \end{bmatrix}$$

After estimating the multivariate AR(1) parameters, Eq. (5.74) could be used to generate synthetic data for the stations.

5.13.2 Disaggregation Modeling

Disaggregation modeling is a process by which time series are generated depending on a time series already available. Disaggregation models allow us to preserve statistics of time series at more than one level. The levels may be represented in both time and space. Disaggregation models have two major benefits. First, these techniques reduce the number of parameters in comparison with the other data-generation models. Second, disaggregation allows for increased flexibility in the methods used for generation (Salas et al., 1988).

The two basic forms of disaggregation are temporal and spatial models. A temporal example is the disaggregation of an annual time series into a seasonal time series. An example of spatial disaggregation is the disaggregation of total natural flow of a river basin into individual tributary flows.

The general disaggregation is a linear dependence model as follows:

$$Y = AX + B\varepsilon \tag{5.86}$$

where Y is the subseries or dependent series generated based on the key (independent) time series X; ε represents the current value from a completely random series; and A and B are the model parameters. In general, the above variables and parameters are represented in the form of vectors and matrices.

Two other disaggregation models are the extended and condensed models. The extended model developed by Mejia and Rousselle (1976) is an extension of the basic temporal model. An additional term is added to the model to consider seasonal covariance between seasons of the current year and seasons of the past year. The model takes the form:

$$Y = AX + B\varepsilon + CZ \tag{5.87}$$

where Z is a column matrix containing seasonal values from the previous years, and C is an additional parameter matrix.

Lane (1979) has developed a condensed model that can be written as:

$$Y_\tau = A_\tau X + B_\tau \varepsilon + C_\tau Y_{\tau-1} \tag{5.88}$$

The subscript τ denotes the current season being generated. Parameters for the linear, extended, and condensed disaggregation models can be estimated as follows (Salas et al., 1988):

For the linear model:

$$\hat{A} = S_{YX} S_{XX}^{-1} \tag{5.89}$$

$$\hat{B}\hat{B}^T = S_{YY} - \hat{A} \cdot S_{XY} \tag{5.90}$$

where $\hat{A}$ and $\hat{B}$ are the estimated parameters. In a temporal disaggregation model, S_{YX} is the matrix of covariances between the seasonal and annual series, S_{XX} is the matrix of covariances among the annual series, and S_{YY} is the matrix of covariances among the seasonal series.

For the extended model, the parameters of the model are estimated as follows:

$$\hat{A} = \left(S_{YX} - S_{YZ} S_{ZZ}^{-1} S_{ZX}\right)\left(S_{XX} - S_{XZ} S_{ZZ}^{-1} S_{ZX}\right)^{-1} \tag{5.91}$$

$$\hat{C} = \left(S_{YZ} - \hat{A} S_{XZ}\right) S_{ZZ}^{-1} \tag{5.92}$$

$$\hat{B}\hat{B}^T = \left(S_{YY} - \hat{A} S_{XX} \hat{A}^T - \hat{A} S_{XZ} \hat{C}^T - \hat{C} S_{ZX} \hat{A}^T - \hat{C} S_{ZZ} \hat{C}^T\right) \tag{5.93}$$

Or, equivalently,

$$\hat{B}\hat{B}^T = S_{YY} - \hat{A} S_{XY} - \hat{C} S_{ZY} \tag{5.94}$$

The notation is the same as for the basic linear model. Z consists of the lagged partial seasonal series, which are used in addition to those used in the linear basic model. S_{ZX} and S_{ZY} are the matrix of covariances between the lagged partial seasonal series and the annual and seasonal series, respectively.

The parameters of the condensed model are estimated as:

$$\begin{aligned} \hat{A}_\tau = &\left[S_{YX}(\tau,\tau) - S_{YY}(\tau-1,\tau) S_{YY}^{-1}(\tau-1,\tau-1) S_{YX}(\tau-1,\tau)\right] \times \\ &\left[S_{XX}(\tau,\tau) - S_{XY}(\tau,\tau-1) S_{YY}^{-1}(\tau-1,\tau-1) S_{YX}(\tau-1,\tau)\right]^{-1} \end{aligned} \tag{5.95}$$

$$\hat{C}_\tau = S_{YY}(\tau,\tau-1) - \hat{A}_\tau S_{XY}(\tau,\tau-1) S_{YY}^{-1}(\tau,\tau-1) \tag{5.96}$$

and

$$\hat{B}_\tau \hat{B}_\tau^{\,T} = S_{YY}(\tau,\tau) - \hat{A}_\tau S_{XX}(\tau,\tau) - \hat{C}_\tau S_{YY}(\tau-1,\tau) \tag{5.97}$$

The condensed model has one equation for each season. The current season (τ) is denoted by τ. As an example, for covariance matrices, $S_{XY}(\tau,\tau-1)$ represents the covariance matrix between the annual series associated with the current season (τ) and seasonal values associated with the previous season (τ-1). See Salas et al. (1988) for more details.

5.14 PROBLEMS

5.1 The following table shows the annual rainfall at a rain gauge. Test the stationarity of the time series in its mean. Define a trend model and the residuals of the series.

Time Sequences	Precipitation Data (mm)	Time Sequences	Precipitation Data (mm)
1	200	11	172
2	212	12	180
3	302	13	185
4	212	14	190
5	192	15	302
6	186	16	215
7	148	17	218
8	312	18	158
9	220	19	142
10	195	20	175

5.2 A series of residuals of a model fitted to precipitation data in a station in 20 time intervals is given in the following table. Use the turning point test and comment on the randomness of the residuals.

Month	Residual	Month	Residual
1	0.87	11	0.78
2	0.28	12	0.82
3	0.32	13	0.14
4	0.34	14	0.22
5	0.33	15	0.65
6	0.15	16	0.48
7	0.18	17	0.32
8	0.17	18	0.07
9	0.21	19	0.24
10	0.22	20	0.25

5.3 The classified monthly precipitation data measured in a station are divided into eight categories as shown in the following table. Previous studies have shown that the precipitation data in this station are normally distributed. Use the chi-square test with a 5% significance level and comment on the selected distribution for the sample data.

Interval (i)	Precipitation Depth Classes (mm)	Actual Frequency (f_i)
1	$0 \le P < 5$	27
2	$5 \le P < 10$	24
3	$10 \le P < 15$	9
4	$15 \le P < 20$	8
5	$20 \le P < 25$	5
6	$25 \le P < 30$	4
7	$30 \le P < 35$	2
8	$35 \le P < 0$	1

5.4 A streamflow time series of a river could be shown in the form of the following equation. Find the extended formulation for this model.

$$(1 - B^{12})Z_t = (1 + 0.2B)(1 - 0.9B^{12})\varepsilon_t$$

5.5 Summarize the main properties of the multivariate and disaggregation models? What is the benefit of these models in comparison to the univariate model?

5.6 Write the mathematical expression of a first-order and second-order nonseasonal ARIMA. Why is the seasonal differencing done? Write the mathematical expression for the second-order seasonal differencing for a monthly time series (assume 12 for the number of seasons).

5.7 An ARMA(1,1) model is fitted to the time series of streamflow at a station with parameters $\varphi = -0.6$ and $\theta = -0.4$. Plot the autocorrelation functions of the model from lag-1 to lag-5.

5.8 Estimate the autocorrelation and partial autocorrelation coefficients for lag-1 to lag-5 for the streamflow data shown in the following table. Assume that the data is normal. Estimate the parameters of AR(1) and AR(2) models and standard deviation of the residuals for both models.

5.9 For the data of Problem 5.8, estimate the parameters of the MA(1), MA(2), and ARMA(1,1). Use AR(1) and ARMA(1,1) models and generate 5 synthetic data.

Discharge (m^3/sec)	Year	Discharge (m^3/sec)	Year
14.6	18	59.8	1
44.2	19	63.3	2
131.0	20	57.7	3
73.3	21	64.0	4
46.6	22	63.5	5
39.4	23	38.0	6
41.2	24	74.6	7
41.3	25	13.1	8
62.0	26	37.6	9
90.0	27	67.5	10
50.6	28	21.7	11
41.9	29	37.0	12
35.0	30	93.5	13
16.5	31	58.5	14
35.3	32	63.3	15
30.0	33	74.4	16
52.6	34	34.2	17

References

Akaike, H., A New Look at the Statistical Model Identification, IEEE Trans. on Automatic Control, AC-19, 6, 716–723, 1974.

Anderson, T. W., Distribution of the Serial Correlation Coefficients, *Annals of Math. Stat.*, 13(1), 1–13, 1942.

Box, G. E. P. and Cox, D. R., An analysis of transformation, *J. Roy. Stat. Soc.*, B26, 211–252, 1964.

Box, G. E. P. and Jenkins, G. M., *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1970.

Box, G. E. P. and Pierce, D. A. Distribution of the residual autocorrelations in autoregressive-integrated moving average time series models, *J. Am. Stat. Assoc*., 65, 1509–1526, 1970.

Brockwell, P. J. and Davis, R. A., *Time Series Theory and Methods*, Springer-Verlag, New York, 1987.

Delleur, J. W., and Karamouz, M., Uncertainty in reservoir operation, *Optimal Allocation of Water Resources*, International Association of Hydrological Sciences, Publications No. 135, Exeter, U.K., 1982.

Fiering, M. B., Multivariate techniques for synthetic hydrology, *J. Hydraulic Division*, ASCE, 90, HY5, 43–60, 1964.

Granger, C. W. J., and Anderson, A. P., Non-linear time series modeling, *Applied Time Series Analysis*, D. F. Findley (ed.), Academic Press, New York, 1978.

Hipel, K. W. and McLeod, A. I., Advances in Box-Jenkins modeling, 1. Model construction, *Water Resources Res.*, 13(3), 567–575.

Lane, W. L., *Applied Stochastic Techniques (LAST Computer Package) User Manual*, Division of Planning Technical Services, Bureau of Reclamation, Denver, CO, 1979.

Lapin L. L., *Probability and Statistics for Modern Engineering*, PWS-Kent Publishing, Boston, MA, 1990.

Ljung, G., and Box G. E. P., On a measure of lack of fit in time series models, *Biometrika*, (65), 297–303, 1978.

McLeod, A. I., and Li, W. K., Diagnostic Checking of ARMA Time Series Models Using Squared-Residual Autocorrelations, *J. Time Series Anal*., 4, 269–273, 1983.

Matalas, N. C., Mathematical assessment of synthetic hydrology, *J. Water Resources Res.*, 3(4), 937–945, 1967.

Mejia, J. M., *On the Generation of Multivariate Sequences Exhibiting the Hurst Phenomenon*, Fort Collins, CO, 1971.

Mejia, J. M. and Rousselle, J., Disaggregation models in hydrology revisited, *J. Water Resources Res.*, 12(2), 185–186, 1976.

O'Connell, P. E., Stochastic Modeling of Long-Term Persistence in Streamflow Sequences, Ph.D. thesis, Imperial College, University of London, 1974.

Salas, J. D. and Pegram, G. G. S., A seasonal multivariate multilag autoregressive model in hydrology, in *Modeling Hydrologic Processes*, Morel-Seytoux, H., Salas, J. D., Sanders, T. G., and Smith, R. E., Eds., Water Resources Publications, Fort Collins, CO, 1978.

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L., A*pplied Modeling of Hydrological Time Series*, Water Resources Publication, Fort Collins, CO, 1988.

Shahin, M., Van Oorschot, H. J. K., and De Lang, S. J., *Statistical Analysis in Water Resources Engineering*, A. A. Balkema, Rotterdam, the Netherlands, 1993.

Snedecor, G.W. and W.G. Cochran, *Statistical Methods*, 6th edition, Iowa State University Press, Ames., Iowa, 1967.

Tao, P. C. and Delleur, J. W., Seasonal and nonseasonal ARMA models in hydrology, ASCE *J. Hydraulic Division*, 102, HY10, 1541–1559, 1976.

Valencia, D. and Schaake, J. C. Jr., Disaggregation process in stochastic hydrology, *J. Water Resources Res.*, 9(3), 580–585, 1973.

Yevjevich, V., *Stochastic Process in Hydrology*, Water Resources Publications, Fort Collins, CO, 1972.

APPENDIX A

Chi-Square Distribution (The following table provides the values of χ^2_α that correspond to a given upper-tail area α and a specified number of degrees of freedom.)

TABLE A.1

Degrees of Freedom	Upper-Tail Area α						
	0.99	**0.98**	**0.95**	**0.90**	**0.80**	**0.70**	**0.50**
1	0.000157	0.000628	0.00393	0.0158	0.0642	0.148	0.455
2	0.0201	0.0404	0.103	0.211	0.446	0.713	1.386
3	0.115	0.185	0.352	0.584	1.005	1.424	2.366
4	0.297	0.429	0.711	1.064	1.649	2.195	3.357
5	0.554	0.752	1.145	1.610	2.343	3.000	4.351
6	0.872	1.134	1.635	2.204	3.070	3.828	5.348
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337
26	12.189	13.409	15.379	17.292	19.820	21.792	25.336
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336
30	14.953	16.306	18.493	20.599	23.364	25.508	26.336

(continued)

TABLE A.1 (CONTINUED)

Degrees of Freedom	Upper-Tail Area α						
	0.30	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
1	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	3.665	4.642	6.251	7.815	9.837	11.345	16.268
4	4.878	5.989	7.779	9.488	11.668	13.277	18.465
5	6.064	7.289	9.236	11.070	13.388	15.086	20.517
6	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	33.530	36.250	40.256	43.773	47.962	50.892	59.703

APPENDIX B

Critical Values of D for Kolmogorov Smirnov Maximum Deviation Test for Goodness of Fit (The following table provides the critical values, D_α, corresponding to an upper-tail probability, α of the test statistic D.)

TABLE B.1

n	α = 0.005	α = 0.01	α = 0.025	α = 0.05	α = 0.1
1	0.99500	0.99000	0.97500	0.95000	0.90000
2	0.92929	0.90000	0.84189	0.77639	0.68377
3	0.82900	0.78456	0.70760	0.63604	0.56481
4	0.73424	0.68887	0.62394	0.56522	0.49265
5	0.66853	0.62718	0.56328	0.50945	0.44698
6	0.61661	0.57741	0.51926	0.46799	0.41037
7	0.57581	0.53844	0.48342	0.43607	0.38148
8	0.54179	0.50654	0.45427	0.40962	0.35831
9	0.51332	0.47960	0.43001	0.38746	0.33910
10	0.48893	0.45662	0.40925	0.36866	0.32260
11	0.46770	0.43670	0.39122	0.35242	0.30829
12	0.44905	0.41918	0.37543	0.33815	0.29577
13	0.43247	0.40362	0.36143	0.32549	0.28470
14	0.41762	0.38970	0.34890	0.31417	0.27481
15	0.40420	0.37713	0.33760	0.30397	0.26588
16	0.39201	0.36571	0.32733	0.29472	0.25778
17	0.38086	0.35528	0.31796	0.28627	0.25039
18	0.37062	0.34569	0.30936	0.27851	0.24360
19	0.36117	0.33685	0.30143	0.27136	0.23735
20	0.35241	0.32866	0.29408	0.26473	0.23156
21	0.34427	0.32104	0.28724	0.25858	0.22617
22	0.33666	0.31394	0.28087	0.25283	0.22115
23	0.32954	0.30728	0.27490	0.24746	0.21645
24	0.32286	0.30104	0.26931	0.24242	0.21205
25	0.31657	0.29516	0.26404	0.23768	0.20790
26	0.31064	0.28962	0.25907	0.23320	0.20399
27	0.30502	0.28438	0.25438	0.22898	0.20030
28	0.29971	0.27942	0.24993	0.22497	0.19680
29	0.29466	0.27471	0.24571	0.22117	0.19348
30	0.28987	0.27023	0.24170	0.21756	0.19032
31	0.28530	0.26596	0.23788	0.21412	0.18732
32	0.28094	0.26189	0.23424	0.21085	0.18445
33	0.27677	0.25801	0.23076	0.20771	0.18171
34	0.27279	0.25429	0.22743	0.20472	0.17909
35	0.26897	0.25073	0.22425	0.20185	0.17659
36	0.26532	0.24732	0.22119	0.19910	0.17418
37	0.26180	0.24404	0.21826	0.19646	0.17188
38	0.25843	0.24089	0.21544	0.19392	0.16966

(continued)

TABLE B.1 (CONTINUED)

n	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.1$
39	0.25518	0.23786	0.21273	0.19148	0.16753
40	0.25205	0.23494	0.21012	0.18913	0.16547
41	0.24904	0.23213	0.20760	0.18687	0.16349
42	0.24613	0.22941	0.20517	0.18468	0.16158
43	0.24332	0.22679	0.20283	0.18257	0.15974
44	0.24060	0.22426	0.20056	0.18053	0.15796
45	0.23798	0.22181	0.19837	0.17856	0.15623
46	0.23544	0.21944	0.19625	0.17665	0.15457
47	0.23298	0.21715	0.19420	0.17481	0.15295
48	0.23059	0.21493	0.19221	0.17302	0.15139
49	0.22828	0.21277	0.19028	0.17128	0.14987
50	0.22604	0.21068	0.18841	0.16959	0.14840
51	0.22386	0.20864	0.18659	0.16796	0.14697
52	0.22174	0.20667	0.18482	0.16637	0.14558
53	0.21968	0.20475	0.18311	0.16483	0.14423
54	0.21768	0.20289	0.18144	0.16332	0.14292
55	0.21574	0.20107	0.17981	0.16186	0.14164
56	0.21384	0.19930	0.17823	0.16044	0.14040
57	0.21199	0.19758	0.17669	0.15906	0.13919
58	0.21019	0.19590	0.17519	0.15771	0.13801
59	0.20844	0.19427	0.17373	0.15639	0.13686
60	0.20673	0.19267	0.17231	0.15511	0.13573
61	0.20506	0.19112	0.17091	0.15385	0.13464
62	0.20343	0.18960	0.16956	0.15263	0.13357
63	0.20184	0.18812	0.16823	0.15144	0.13253
64	0.20029	0.18667	0.16693	0.15027	0.13151
65	0.19877	0.18525	0.16567	0.14913	0.13052
66	0.19729	0.18387	0.16443	0.14802	0.12954
67	0.19584	0.18252	0.16322	0.14693	0.12859
68	0.19442	0.18119	0.16204	0.14587	0.12766
69	0.19303	0.17990	0.16088	0.14483	0.12675
70	0.19167	0.17863	0.15975	0.14381	0.12586
71	0.19034	0.17739	0.15864	0.14281	0.12499
72	0.18903	0.17618	0.15755	0.14183	0.12413
73	0.18776	0.17498	0.15649	0.14087	0.12329
74	0.18650	0.17382	0.15544	0.13993	0.12247
75	0.18528	0.17268	0.15442	0.13901	0.12167
76	0.18408	0.17155	0.15342	0.13811	0.12088
77	0.18290	0.17045	0.15244	0.13723	0.12011
78	0.18174	0.16938	0.15147	0.13636	0.11935
79	0.18060	0.16832	0.15052	0.13551	0.11860
80	0.17949	0.16728	0.14960	0.13467	0.11787
81	0.17840	0.16626	0.14868	0.13385	0.11716
82	0.17732	0.16526	0.14779	0.13305	0.11645

(continued)

TABLE B.1 (CONTINUED)

n	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.1$
83	0.17627	0.16428	0.14691	0.13226	0.11576
84	0.17523	0.16331	0.14605	0.13148	0.11508
85	0.17421	0.16236	0.14520	0.13072	0.11442
86	0.17321	0.16143	0.14437	0.12997	0.11376
87	0.17223	0.16051	0.14355	0.12923	0.11311
88	0.17126	0.15961	0.14274	0.12850	0.11248
89	0.17031	0.15873	0.14195	0.12779	0.11186
90	0.16938	0.15786	0.14117	0.12709	0.11125
91	0.16846	0.15700	0.14040	0.12640	0.11064
92	0.16755	0.15616	0.13965	0.12572	0.11005
93	0.16666	0.15533	0.13891	0.12506	0.10947
94	0.16579	0.15451	0.13818	0.12440	0.10889
95	0.16493	0.15371	0.13746	0.12375	0.10833
96	0.16408	0.15291	0.13675	0.12312	0.10777
97	0.16324	0.15214	0.13606	0.12249	0.10722
98	0.16242	0.15137	0.13537	0.12187	0.10668
99	0.16161	0.15061	0.13469	0.12126	0.10615
100	0.16081	0.14987	0.13403	0.12067	0.10563

6 River Basin Modeling

6.1 INTRODUCTION

This chapter discusses elements of river basin modeling, including data and information processing, objectives, and constraints identification, and details of modeling process. The objectives of river basin modeling that are discussed include water quality and quantity management and flood control. Also presented in this chapter are managerial tools, such as simulation and optimization models, and a number of computer packages that have been widely used in various water resources development and management projects.

6.2 WATERSHED: DEFINITION AND CHARACTERISTICS

The concept of watershed is the foundation of all hydrologic planning and design. An understanding of watersheds — formed by rivers and the land area they drain — is basic for understanding the major sources of freshwater all over the world. Sound water resources management must be undertaken at a whole-watershed level, rather than just at a local level. As shown in Figure 6.1, many major watersheds cross national boundaries (World Resources Institute, 1998).

A watershed is also referred to as a *catchment basin*, which can be regarded as a system as it has a well-defined boundary and its elements show clear relationships both structurally (in terms of morphology) and functionally (by virtue of the flow of matter and energy). The inputs and outputs across the system boundary can also be clearly distinguished (Figure 6.2). The areal limits of the river system are set by a major watershed, while minor watersheds define subcatchments within it (Figure 6.3). The boundary of a watershed is defined by all points that will shed water to the outlet; it is only necessary to determine which points in a region contribute water to the outlet, and the most extreme points represent the watershed boundary. In the next sections some characteristics of a watershed are defined.

6.3 WATERSHED GEOMORPHOLOGY

6.3.1 Drainage Area

The drainage area is an important watershed characteristic for hydrologic planning and design because it contains the volume of water generated from rainfall. Determining the drainage area requires delineation of the watershed boundary. The Geographic Information System (GIS) is a powerful tool for this purpose.

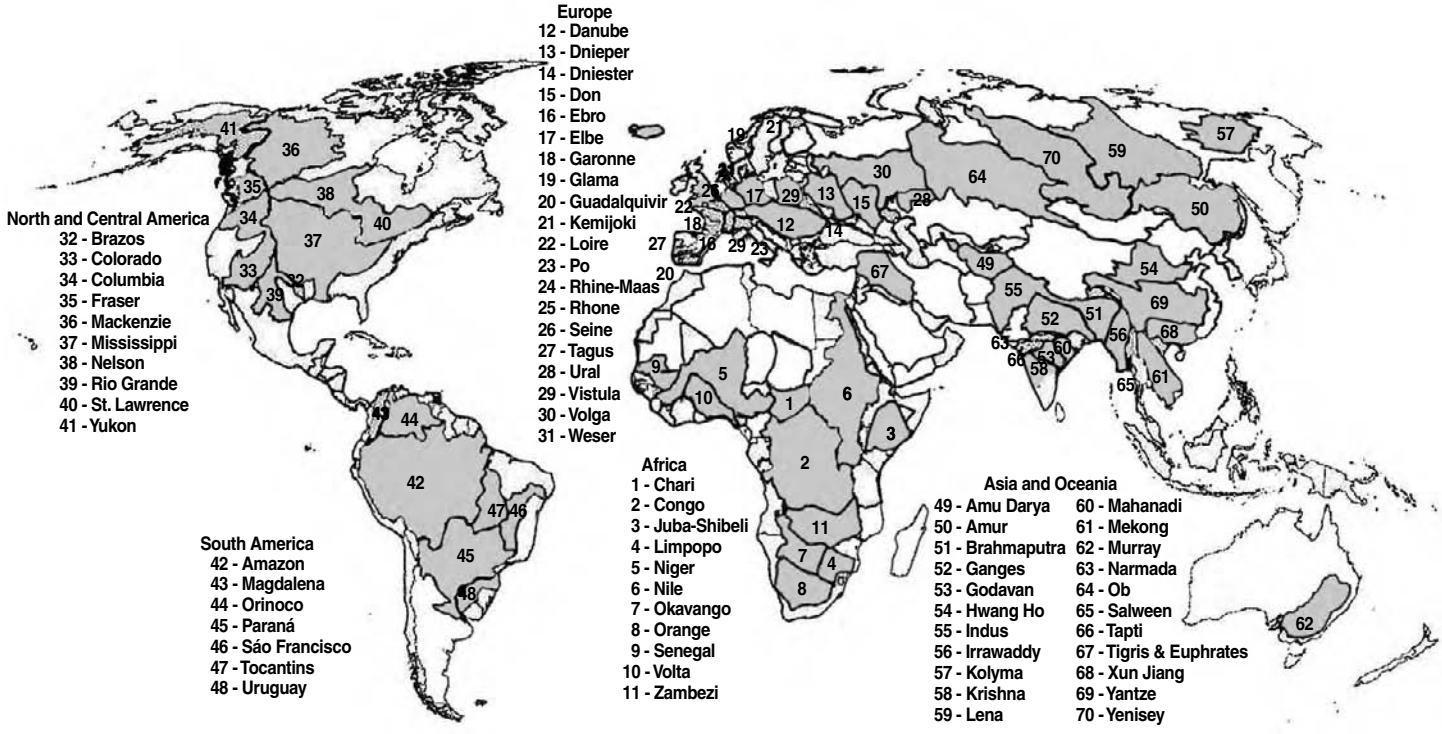


FIGURE 6.1 Major watersheds across national boundaries. (From World Resources Institute, *World Resources: A Guide to the Global Environment*, a joint publication by The World Resources Institute, The United Nations Environment Programme, The United Nations Development Programme, and The World Bank, Oxford University Press, London, 1998.)

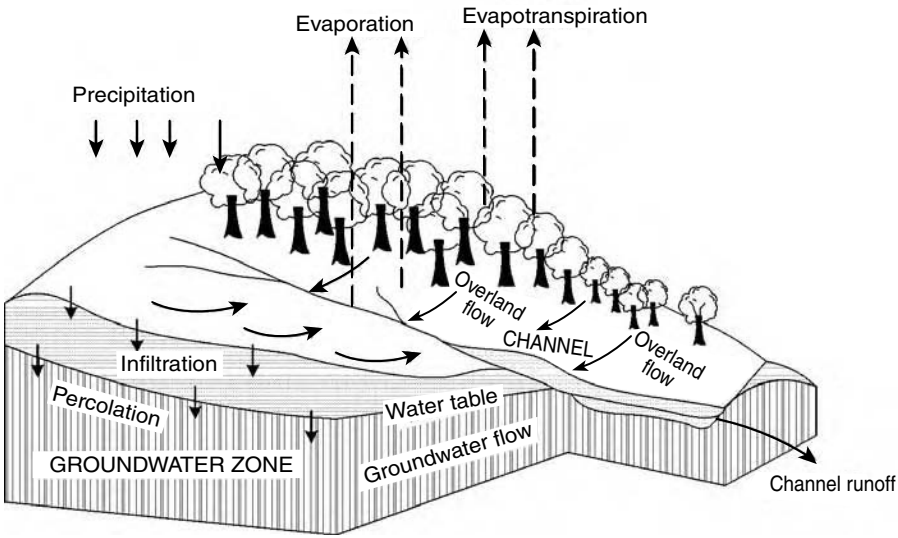


FIGURE 6.2 Catchment basin system.

6.3.2 Watershed Length

Watershed length is defined as the distance measured from the watershed outlet to the farthest point on the basin divide. The length is usually a measure of the travel time of water through the watershed.

6.3.3 Watershed Slope

Watershed slope is an important factor in the momentum of runoff and flood magnitude. It is the rate of change of elevation with respect to distance along the principal flow path. According to the simple classification scheme of White et al. (1992), based on the surface gradient and slope plan form, any catchment basin consists of no more than five fundamental types of slopes. White et al. (1992) classified these slopes into two groups: flat (or nearly so) and valley (Figure 6.4). Where it is required to subdivide the watershed, it is necessary to compute the slopes of each subarea. For this purpose, the principal flow path for that subarea must be delineated. The path is divided into reaches, and the river slope in different reaches may be required for computing the reach travel time.

6.3.4 Watershed Shape

Watersheds have an infinite variety of shapes, and parameters that reflect the basin shape are used in hydrologic planning and design. Some typical parameters that have been developed to reflect watershed shape are as follows:

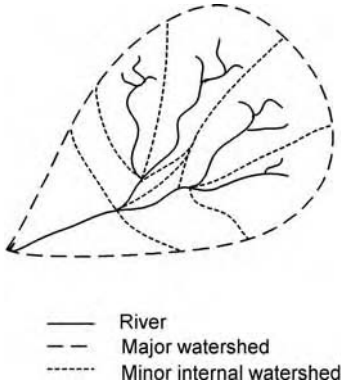


FIGURE 6.3 Delimitation of catchment watershed.

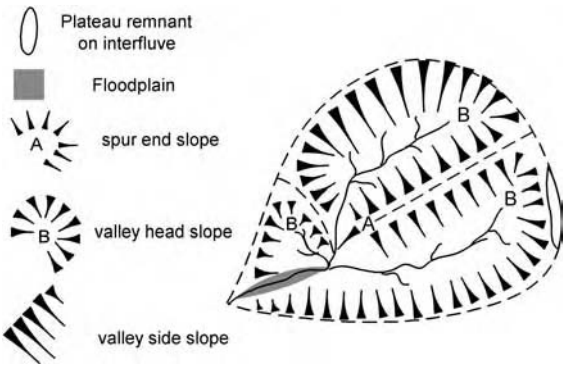


FIGURE 6.4 Morphological components of the catchment basin.

- Length to the center of area (L_{ca}) is the distance (measured in kilometers) along the main channel between the basin outlet and the point on the main channel opposite the center of the area.
- Shape factor (L_l) is estimated by the following equation:

$$L_l = \left(L \cdot L_{ca} / 2.58\right)^{0.3} \tag{6.1}$$

where L is the watershed length in kilometers. L_l shows the elongation of a watershed.

- Circularity ratio (R_c) is estimated by the following equation:

$$R_c = A \big/ A_0 \tag{6.2}$$

where A is area of watershed, and A_0 is the area of a circle with perimeter equal to the perimeter of the basin.

6.4 SOIL CHARACTERISTICS

The three soil phase systems are solid, liquid, and gas. The liquid and gas phases are represented by the soil water and air, respectively, that occupy the pores and voids. However solid, liquid, and gas phases do not form a random mosaic in these three dimensions but are more or less organized to impart a definite vertical and lateral structure to the system. The properties of the soil solids and of the voids and their associated soil water and soil air vary vertically and laterally. The vertical component of the soil assembly is referred to as the *soil profile* and the soil layers as its *horizon*. Soil horizons are referred to as follows (McCuen, 1998):

O horizon — Surface litter consisting primarily of organic matter
A horizon — Top soil consisting of humus and inorganic minerals
E horizon — Leaching zone where percolating water dissolves water-soluble matter
B horizon — Subsoil below the A or E horizon that contains minerals and humic compounds
C horizon — A zone consisting primarily of under-composed mineral particles and rock fragments
R horizon — An impermeable layer of bedrock

6.4.1 Soil Texture

Soil texture refers to the size of mineral particles in different size categories. On the base of mean diameter (d), particles of soil can be separated into four classes:

1. Gravel, $d \leq 2$ mm
2. Sand, $0.02 \leq d \leq 2$ mm
3. Silt, $0.002 \leq d \leq 0.02$ mm
4. Clay, $d \leq 0.002$ mm

Water-holding characteristics and the infiltration capacity of the soil can be determined by the soil texture. The capacity of soil to pass and store infiltrating water increases with the size of the soil particles as a result of the increase in pore space.

6.4.2 Soil Structure

The kind, size, and distribution of soil aggregates and soil voids and pores determine the structure or fabric of the soil. Soil structure refers to the tendency of the soil particles to aggregate into lump and clods. The hydrologic response of the watershed, soil moisture, and water movement through the column of soil can be affected by the structure of the soil.

6.4.3 Soil Moisture

The water in the unsaturated zone of soil is referred to as *soil moisture* and is acted upon by gravitation, surface tension, and molecular forces. Soil moisture is held as

capillary water in the smaller pore spaces of the soil or as hygroscopic water absorbed on the surface of soil particles. Capillary water can be removed from soil by applying forces sufficient to overcome the capillary forces. *Field capacity* is the moisture content of the soil after drainage. Drainage of water will continue for some time after a soil is saturated, so field capacity must be defined in terms of a specific drainage period. The *permanent wilting point* is a moisture content at which plants can no longer extract sufficient water from soil for growth. Water below this level of soil moisture is referred to as *unavailable water*, and the only moisture presented in the soil is the thin film or layer of moisture attached to the soil particles, which is referred to as *hygroscopic moisture*. *Available water* is the moisture content of the soil that exists between field capacity and the permanent wilting point. An efficient irrigation procedure is to apply water when the moisture content of the soil approaches the wilting point in an amount sufficient to raise the soil moisture to the field capacity within the root zone.

6.5 DATA REQUIREMENTS

6.5.1 Hydrologic and Climatologic Data

Water resources planning and management requires basin climatic and hydrologic data. The movement of water through air, sea, lakes, rivers, land, soil, glaciers, and living organisms forms the hydrologic cycle. The atmospheric part of the cycle begins when evaporation occurs off of open water surfaces, as well as off of ice, plants, soil, and all other surfaces recently wetted by precipitation.

A portion of rainfall is intercepted by the leaves and stems of vegetation. Interception, along with both depression storage and soil moisture, constitutes *basin recharge* — that portion of precipitation that does not contribute to surface runoff or groundwater. Depression storage includes the water retained as puddles in surface depressions. Some of the rainwater or melted snow overflows as *surface runoff* and *overland flow*. Other water may infiltrate into the soil and flow laterally in the soil surface to a stream channel as *interflow*. Overland flow and interflow are usually grouped together as *direct runoff*.

Data gathered from gauging stations can be used to determine stream flows and estimate interflows. *Infiltration* is the movement of water through the soil surface and into the soil. The infiltration capacity of a soil at any time is the maximum rate at which water will enter the soil. The infiltration rate is the rate at which water actually enters the soil during a storm. Soil texture and structure, moisture content, land cover and use, and rainfall intensity are the types of data that can be used to determine water losses due to infiltration.

Climate interacts with soil, rock, plants, animals, surface water, and ice. Within the hydrologic cycle, water is stored at a number of locations. The major part of it is stored in the atmosphere, land, oceans, and polar ice caps. Transfers between these locations are carried out by the processes of evaporation, condensation, precipitation, runoff, freezing, and melting. The greatest exchanges are those between oceans and atmosphere. *Evaporation* is the process by which liquid water is changed into its gaseous phase (water vapor). The most important climatic data necessary to estimate

the rate of evaporation are the heat energy supply, humidity of the air, characteristics of the airflow across the surface, and nature of the evaporating surface.

Vegetation provides an extra pathway for water to travel from the ground surface into the atmosphere. Water vapor is diffused through pores on the leaf surface and is transferred into the atmosphere. This transfer of water is referred as *transpiration* and accounts for a large proportion of moisture losses from vegetated surfaces. The rate of transpiration loss is governed by two types of factors: extrinsic factors affecting the rate of evaporation from the water surface and the factors intrinsic to the plant. For example, transpiration loss from most plants will be large when the atmosphere is relatively dry. In most plants, leaf stomata open under light conditions and close at night, thereby introducing a clear diurnal variation in transpiration losses. Therefore, a significant amount of climatic and hydrologic data are required for estimating of transpiration losses, the most important of which are temperature, available water, land cover, wind condition, and geographic latitude. In modeling river basins, a correct estimate of the water budget and the spatiotemporal distribution of data is a key factor. Data accuracy and considering the physical and hydrological characteristics of the watershed is essential to integrated water resources planning of that system.

6.5.2 Physical Characteristics

Watershed and river geomorphology is a basic piece of data necessary for river basin planning and management. These characteristics are important for determination of the volume of water that can be generated from precipitation. As discussed previously, drainage area, watershed length, watershed slope, watershed shape, surface roughness, soil characteristics, and land cover of the river basin are the most important physical characteristics of watershed that should be estimated before undertaking the river basin planning process.

In addition to watershed characteristics, river geomorphology is frequently used in hydrologic computations. The major physical characteristics of rivers or watershed channels are:

1. Length
2. Slope
3. Cross-section
4. Roughness
5. Form

River length and *slope* can be described by several computational schemes; the most common are described here.

The river (channel) *length* is the distance between the river basin outlet and the end of the river along the main channel (L_c). The river (channel) *slope* is the difference in elevation between the upper and lower elevation of river (ΔE_c) over the length (L_c).

$$Sc = \frac{\Delta E_c}{L_c} \tag{6.3}$$

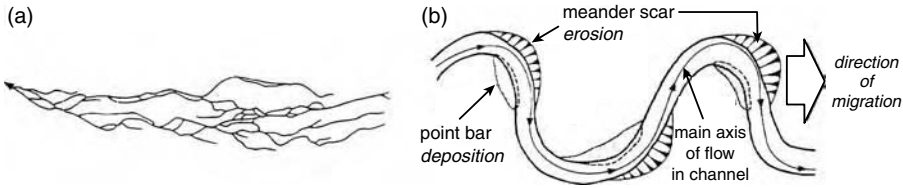


FIGURE 6.5 (a) Braided channel, and (b) geomorphological features of a meandering channel.

In most cases, the river slope is not uniform and a weighted slope can be used that provides better results:

$$S_i = \left(\frac{n}{k}\right)^2 \tag{6.4}$$

where n is the number of river or channel segments and k is given by the following equation:

$$k = \sum_{i=1}^{n} \frac{1}{\left(\Delta E_i / L_i\right)^{0.5}} \tag{6.5}$$

where ΔE_i is the difference in elevation between the points defining the upper and lower segment i of the river, and L_i is the length of segment i. The river slope over each segment is considered relatively constant.

Cross-section information, including cross-section area, wetted perimeters, slope average velocity, and roughness, is usually necessary for hydrologic analysis, planning, and management.

Channel roughness affects the hydraulic characteristics of river flow and is required for a number of hydraulic and hydrologic computations. Manning's roughness coefficient is used in a number of methods for estimation of water elevation, velocity of stream flow, and travel time of runoff, thus channel roughness should be estimated as an important input.

The *channel form* is another important physical characteristic of rivers. The river discharge can increase in the downstream direction as the watershed area increases, thus the channel form changes, becoming deeper and wider. A modeled channel form can demonstrate how the river adjusts to input of sediment and discharge. Channel forms can be classified as straight, sinuous, meandering, and braided channels (see Figure 6.5) (White et al., 1992).

6.5.3 Hydrologic Characteristics

Hydrologic characteristics of watersheds reflect the volume of discharge hydrograph produced by a specific rainfall *hyetograph*. The following parameters, which are mainly used in rainfall-runoff simulation models (such as HEC-1), show the timing of runoff and the shape and the volume of the streamflow hydrograph:

- *Time of concentration* (t_c): Time for a wave to propagate from the most distant point in the watershed to the outlet
- *Lag time* (t_p): Time from the center of mass of rainfall excess to the peak of the hydrograph that can be estimated using following methods: (1) *Snyder's method* (Snyder, 1938):

$$t_p = 0.74\, C_t \left(L\, L_c\right)^{0.3} \tag{6.6}$$

where t_p is the lag time (hr); L is the length of the main stream from the outlet to the divide (km); L_c is a watershed shape parameter, which is the length measured along the main channel from the watershed outlet to a point on the main channel that is perpendicular to the center of the area of the watershed (km); and C_t is a watershed storage coefficient, usually ranging from 1.8 to 2.2. (2) *Linsley's method* (Linsley et al., 1982):

$$t_p = 0.72\, C_l \left(\frac{L\, L_c}{\sqrt{S}}\right)^{0.33} \tag{6.7}$$

where t_p is the lag time (hr); L is the length of main stream from the outlet to the divide (km); L_c is a watershed shape parameter, which is the length measured along the main channel from the watershed outlet to a point on the main channel that is perpendicular to the center of the area of the watershed (km); S is the basin slope; and C_l is a watershed storage coefficient, usually ranging from 1.2 for mountainous basins to 0.35 for valleys.

- *Storage coefficient* (R): Shows the storage capacity of the drainage system of the watershed. Linsley et al. (1982) developed the following equation for estimating storage coefficient (R):

$$R = 0.49\, C_r \left(\frac{L\sqrt{A}}{S}\right) \tag{6.8}$$

where A is the watershed area, and C_r is an empirical coefficient that depends on climatic and physical characteristics of the watershed, usually varying from 6 to 8%.

A correct estimate of these parameters will insure the effectiveness of flood warning systems as a major component of an overall watershed management system.

Example 6.1

The area of a basin is estimated to be 500 km^2. If L = 25 km and L_c = 10 km. Estimate the lag-time (t_p) using the Snyder and Linsley methods. Assume that C_t = 1.8, S = 0.02, C_l = 0.8, and C_r = 0.07.

Solution: Using Eqs. (6.6) and (6.7), it can be written that:

$$t_P = 0.74 \times 1.8(25 \times 10)0.3 = 6.98 \text{ hr}$$

$$t_p = 0.72 \times 0.8\left(\frac{25 \times 10}{\sqrt{0.02}}\right)^{0.33} = 6.79 \text{ hr}$$

The lag time values that are estimated by the two methods are close.

Application of the above example could be used to determine the amount of lead time needed for any flood control measures.

6.5.4 Inflow and Outflow

If precipitation exceeds the infiltration and evaporation rates, water will begin to collect on the surface. If rainfall continues, overland flow will occur on the surface, and overland flows from different places begin to run together. Water soon collects into small channels that lead to streams. Other parts of the streamflow may come from direct precipitation and interactions between ground and surface water.

The river inputs have spatial and temporal variations. The river flow variation in time is recorded by hydrographs at various locations, and the shape and magnitude of the hydrographs are related to many factors associated with the rainfall and river basin characteristics. The storm hydrograph can be separated into direct runoff and base flow. The seepage from groundwater contribution to the base flow can also be determined. The river discharge and elevation vary over time at any section and also in the downstream direction. The flow hydrograph at any point of the river can be determined using a flow routing procedure from one known river hydrograph at one upstream point. Hydrologic routing relates input $I(t)$, output $O(t)$, and river storage $S(t)$ by the continuity equation:

$$\frac{dS(t)}{dt} = I(t) - O(t) \tag{6.9}$$

When storage $S(t)$ is related to input and output by the storage function, the river length should be divided to some extent based on spatial distribution of tributaries, withdrawals, groundwater–river interactions, and physiologic characteristics of the river. The outflow of each reach can be determined from the known input hydrograph, groundwater seepage, and drainage using a flow routing method.

6.5.5 Demand Points

Determination of demand points and the time series of demand at each point is an important step in river basin planning. Water allocation models require demand quantity, along with demand point characteristics such as distance from resources, elevation, and topographic condition. Demand point characteristics can affect allocated water costs and the required equipment and means for water transfer. Other

information, such as water quality and the physiologic condition of the demand point, is also needed.

6.5.6 Control Points

From a hydraulic standpoint, a control section or control point is a point in the river at which the flow changes from subcritical to supercritical, passing through the critical depth. In water resources planning, however, a control point is a section that experiences a considerable change in discharge or a point of the river that has a withdrawal tributary and a monitoring or gauging station.

6.5.7 Critical Points

Critical points require more attention in river basin planning and management. Important structures, protected areas, and demand points that must have water of a known quantity and/or quality at an acceptable reliability level are some examples of critical points. Critical points must be determined before the planning phase because they could have considerable effects on planning alternatives such as for flood control or upstream reservoir operation.

6.5.8 Water Quality Requirements for Various Uses

Qualitative characteristics of water are important for water resources planning and management. Water quality is evaluated in terms of its physical, chemical, and microbiological properties. Quantitative measurement of these characteristics is necessary for determination of water quality. Water quality requirements are determined based on the intended use of the water; for example, water contaminated by chemicals might be harmful for agricultural use but could be suitable for some other uses. Therefore, the water quality requirements for each type of water use should be determined, along with assessment of raw water quality and selection of suitable water treatment processes, if necessary. Water quality is usually evaluated by the degree to which physical, chemical, and biological standards are met for each water use. For example, most standards control the turbidity, odor, color, taste, and microbiological characteristics of drinking water. More information about water quality standards and qualitative management of water resources is presented in Chapter 9.

6.5.9 Return Flows and Drainage

During or after the precipitation, water entering the ground may be stored temporarily in the upper layer of soil; it may flow laterally above the groundwater and reach a river or other surface waters, or it may infiltrate and reach the groundwater. Excess agricultural water use and water that infiltrates from municipal drainage wells can reach surface water as return flows. Sometimes groundwater partially or completely recharges the surface waters resources. For example, between rainstorms, groundwater discharges to streams. During storms, the stream level rises, and water drains from the stream to the groundwater, raising the groundwater table. Therefore, groundwater level and river stages have a dynamic relationship and are generally

related. Installation of piezometers or monitoring wells and the design of a network for groundwater table monitoring can provide information that is necessary to determine groundwater and river interaction.

6.5.10 Sampling Network

Water quantity and quality monitoring is essential to obtain information on the quantity and physical, chemical, and biological characteristics of water resources. The type of information to be gathered is determined based on the monitoring objectives and on the statistical water sampling procedures and guidelines. Figure 6.6 demonstrates the flow of information and functional summary of water quality monitoring systems.

One of the most critical factors in a monitoring network is the location of the monitoring and sampling stations. Data on water quantity, water quality, and groundwater depth are collected. The factors that influence the selection of monitoring and sampling stations are different and are determined based on the objectives and constraints of the monitoring system. For example, in a groundwater monitoring network, cost factors may lead to the decision to incorporate preexisting wells into the network. In addition to site or variable selection, sampling frequency is an important aspect of network design. A high percentage of monitoring network operation costs are related to sampling frequency. In statistical assessment of qualitative and quantitative variables, professional judgment and cost constraints are often considered as the basis for sampling frequencies. After network design, operating procedures need to be specified before sampling or data collection. The procedure includes sample collection, sample and sampling preservation methods, chain of custody, laboratory analysis methods, quality control, and data verification certification guidelines.

6.6 OBJECTIVES

6.6.1 Water Supply

Supplying water of the required quantity and quality is one of the most important objectives of water resources planning. Determination of current water use and prediction of future water demands are necessary for water supply planning and management. Domestic or municipal, industrial, and agricultural uses represent the major demands for water resources. Hydropower generation, recreation, and fishing are other types of water use that do not require direct withdrawal from the resources and are called *instream uses*. Each of these uses has different water quality requirements. Surface and groundwater are two main sources of water supply; in this chapter, the focus is on surface water.

Rivers and lakes are important resources for public water supplies but surface waters are being polluted by industrial and municipal wastes, agricultural return flow, and polluted runoffs. Thus, an expensive treatment process is usually necessary before this water could be consumed.

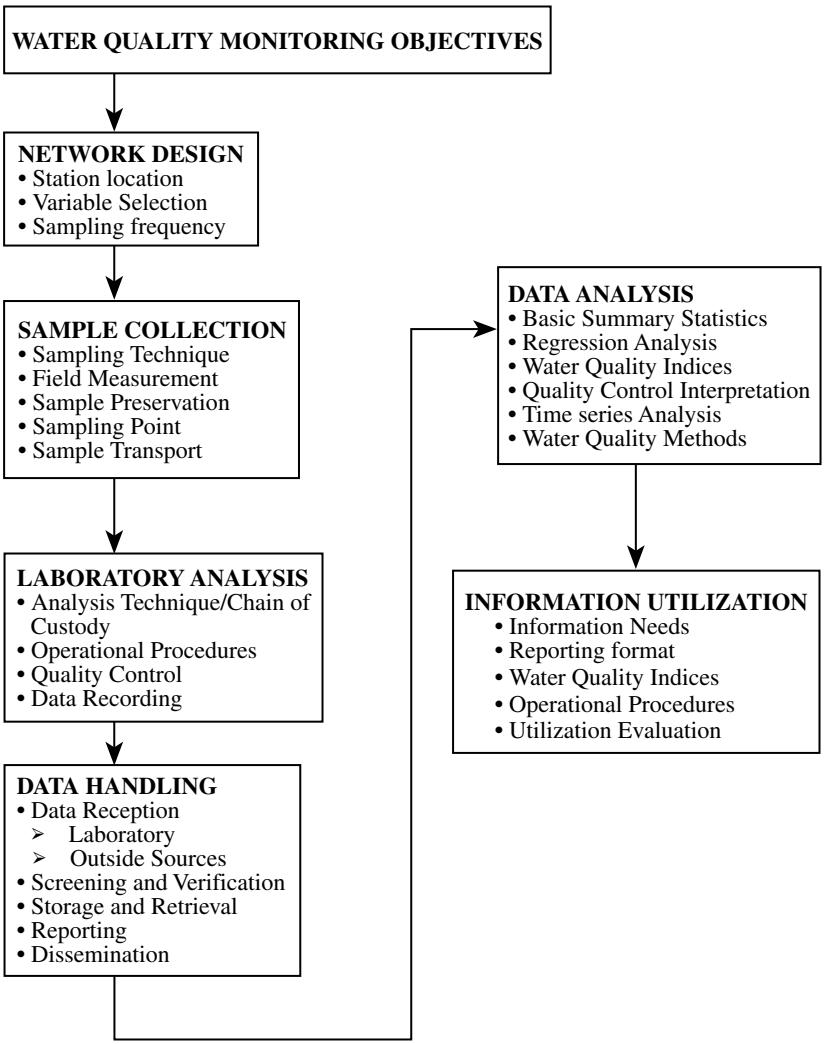


FIGURE 6.6 Flow of information and functional summary of monitoring system activities. (From Sanders, T. G. et al., *Design of Networks for Monitoring Water Quality*, Water Resources Publications, Littleton, CO, 1983. With permission.)

Wastewater reuse is an important alternative for water supply. Wastewater reuse may be more attractive in arid areas where water is scarce and water losses are high. Municipal wastewater is the best resource for water reuse in terms of its quantity and low variability. Advanced wastewater treatment is required for direct municipal reuse. In indirect wastewater reuse, wastewater usually is discharged to the groundwater aquifer and then is withdrawn for potable or other municipal uses. Wastewater recycling and reuse in industries provides savings in energy and costs. Return flows from irrigation are a reusable source for agricultural or industrial uses. When reused

water is allocated for agricultural uses, suitable treatment is required because it is highly contaminated with salt and chemical matters from the use of fertilizers, pesticides, and herbicides. More details about water reuse are presented in Chapter 11.

6.6.2 Flood Control

A *flood* or *flooding* has been defined by the Water Resources Council as a general or temporary condition of partial or complete inundation of normally dry areas by the overflow of inland or tidal waters and/or the unusually rapid accumulation of surface waters (Viessman and Hammer, 1985). During a flood, the overflows from river or channel banks cover the floodplain with lower velocity than in the main river, thus the floodplain can store a high quantity of water and can act as a wider waterway. Floods cause loss of life and property within the river floodplain as well as added costs due to cleanup of the flooded lands.

Flood control and flood damage reduction are important objectives of river basin planning. Flood control and floodplain management require hydrologic and hydraulic analysis of floods. This analysis determines inundated areas, flood elevation, and characteristics of required hydraulic structures for flood control or flood damage reduction. The typical requirements for floodplain analysis and planning include (Hoggan, 1997; Mays et al., 1996):

1. *Floodplain information studies* collect and analyze information regarding specific flood events for various return periods. Floodplain information studies require some hydraulic, morphological, and hydrological data. Satellite information from floodplain feathers can be used for this purpose. As an example, data requirements for flood monitoring in basins larger than 1000 km^2 are summarized in Table 6.1.
2. *Evaluation of future land use alternatives* identifies the potential for flood damage and the environmental impacts of flood events at different frequencies.
3. *Evaluation of flood reduction measures* examines floods with different frequencies to determine expected flood damage reduction associated with specific design flows.
4. *Design studies* determine the scope of structural measures for specific flood events.
5. *Operation studies* develop optimal operating policies for flood control facilities.

Simple steady- or unsteady-state models can be used to determine water surface elevation and streamflow hydrographs at various locations along the river. Water surface elevation and streamflow discharge can be used for potential flood damage estimation. Identification of an alternative or a combination of alternatives for flood damage reduction is the main objective of flood control planning. The best plan can be selected based on its net benefit. The increase in the net income of additional floodplain development and existing floodplain activities is referred to as *location*

TABLE 6.1
Data Requirements for Flood Studies in Basins Larger than 1000 km²

Hydrological Element	Accuracy	Resolution (m)	Frequency
Floodplain area/boundary	±5% area	10	5 years[a]
Flood extent	±5%	100	<= 4 days[b]
Precipitation	±2 mm if <40 mm ±5% if >40 mm	5000	6 hours
Saturated soil area	±5%	100	<= 4 days[b]
Soil moisture profile	±10% field capacity	1000	1 day
Snow-covered area/snow line	±5% area	1000	1 day
Snowpack water equivalent	±2 mm if <2 cm ±10% if >2 cm	1000	1 day
Snowpack free water content	±2 mm if <2 cm ±10% if >2 cm	1000	1 day
Snow surface temperature	±1°C	1000	6 hours

[a] And after each major flood event.
[b] Depends on the degree of flooding; with major floods, daily monitoring would be desirable.

Source: Barrett, E. C. and Curtis, L. F., *Introduction to Environmental Remote Sensing*, Chapman & Hall, New York, 1992.

benefit and *intensification benefit*, respectively. This is computed as the sum of the location benefits and intensification benefits and the flood-inundation-reduction benefit, less the total cost of implementing, operating, maintaining, repairing, replacing, and rehabilitating the system (Mays, 1996).

6.6.2.1 Flood Control Alternatives

Reduction of flood damage is the main objective of flood control planning. Flood control alternatives can be classified into two groups: structural and nonstructural. Structural alternatives represent traditional flood damage reduction by physical means. In other words, the construction of flood control facilities can be referred as structural measures. Important measures for structural reduction of flood damage include:

- *Dams and reservoirs.* Flood control dams may be constructed across the river to store floodwaters and to reduce the magnitude of the flood and the downstream stage of the flood. The stored floodwater can be allocated to different water demands. Flood control reservoirs also can change the hydraulic characteristics and flow regime in downstream of the reservoir.
- *Levees and floodwalls.* Levees and floodwalls are the oldest and commonly used methods of protection against floods. Levees or dikes are constructed parallel to rivers to prevent overflow of floodwater to the floodplain. Floodwalls are usually constructed from concrete and perform the same function as levees. Levees and floodwalls can increase the peak discharge

of floodwater downstream because natural storage in the floodplain is decreased.

- *Flood diversion.* A flood diversion structure or bypass directs excess floodwater from areas with a high potential for flood damage into an area or channel that can carry the flood discharge.
- *Channel modification.* Improvement of the hydraulic capacity of a channel can lower the water stage and increase the carrying capacity. Channel improvement lowers the stage of water for different discharges but reduces natural floodplain storage and results in a higher discharge of flood peak. Straightening of bends or meanders, lining, widening, and deepening are the most important methods of channel modification.
- *Detention measures.* Detention basins are small impoundments with uncontrolled release that can be created by excavations, small dams, or walls. Detention measures store floodwater in the river basin before reaching the main channel. Recharging the groundwater and improving the water quality by reducing sedimentation are among other advantages of onsite detention measures.

Although structural measures have been effective in flood damage reduction, in a comprehensive floodplain plan, both structural and nonstructural alternatives should be evaluated. In a nonstructural approach, flood damage is reduced by developing operational policies for river basin management. Some nonstructural flood damage reduction measures are as follows:

- *Reservoir operation.* In multipurpose reservoirs, temporary storage may be allocated for flood flow or reduction of outflow discharge. Development of suitable operating policies for dam gates and spillways that consider the actual inflow hydrograph and carrying capacity of the downstream river is important for nonstructural flood damage reduction. More details about reservoir operation are provided later.
- *Watershed improvement.* Improving infiltration and groundwater recharge and reducing erosion are the main objectives of watershed modification for flood damage reduction. Improvement of vegetation cover, land use, and terracing management can increase the time of concentration for watersheds and reduce surface soil erosion by reduction of overland flow velocity. Although watershed improvement is a long-term and expensive plan, it can be one of the most effective techniques for protecting of water and soil resources.
- *Floodplain regulation.* The development of land use regulations within main channels and floodplain areas is referred to as floodplain regulation or management. It may consist of many administrative actions that regulate floodplain land use based on potential flood damage. In most regions, government or local agencies restrict floodplain development by floodplain zoning. For example, in many countries, all property owners located within a 100-year floodplain must purchase flood insurance, and the lowest floor or basement of the buildings must be built above the 100-year flood level.

- *Flood forecast and flood warning systems.* Real-time flood forecasting is one of the most effective methods for flood damage reduction. Determination of climatic, hydrologic, and hydraulic and morphological characteristics of the river basin is the first step in flood forecasting. Satellite remote sensing is beginning to contribute useful information with respect to soil moisture, runoff from snowmelt, land cover/use, and rainfall that can be used in flood forecasting and flood warning systems. A flood warning system can provide an opportunity to act before flood occurs. Thus, it can reduce flood damage costs and health hazards. These systems are useful for drainage areas having a concentration time sufficient for controlling floodwater or reducing flood damage. Flood warning preparedness systems send rainfall and runoff information to processing centers when critical thresholds have been reached. In the processing center, downstream discharge and water elevation are calculated, and the time of occurrence and magnitude of downstream flow are estimated and reported. Flood warning preparedness plans, then, can be considered to be a critical and important component of other flood control measures.
- *Flood proofing.* Flood proofing includes all actions by individuals or small groups to modify their sites, structures, and facilities to reduce flood damage. Raising structures, using materials or methods that are less susceptible to flood damage, preventing water from entering structures, and building local levees to restrict and direct floodwater are examples of flood proofing. This approach is an inexpensive method that can be used together with other flood control measures to reduce the risk of flood damage.

6.7 MANAGEMENT TOOLS

6.7.1 Streamflow and Excess Water Estimation

Surface water flow is an important element of water resources planning and design studies for water supply, streamflow forecasting, flood control, irrigation, and reservoir design. Most design problems are caused by designs based on streamflow hydrographs that consider only peak discharge rates. Streamflow and rating curves that correlate discharge with flow rate can be determined in the field using stream velocity measurements and channel geometry. The channel may contain a certain amount of base flow even in the absence of rainfall. Base flow occurs due to water stored in watersheds from past storm events and snowpacks and groundwater contribution and appears as streamflow. Direct runoff from rainfall excess and base flow make up the total hydrograph. The unit hydrograph transforms the rainfall excess into direct runoff.

6.7.1.1 Streamflow Estimation in Homogenous Basins

The selection of appropriate methods for streamflow estimation is based on the characteristics of the river basin. In a drainage area nearly homogenous, the streamflow at each point of the watershed main channel is related to the watershed area

above that point. Therefore, the streamflow at site s can be estimated based on measured flow at gagged point s' (Loucks et al., 1981):

$$Q_t^s = Q_t^{s'} \cdot \left(\frac{A^s}{A^{s'}} \right) \tag{6.10}$$

where A^s and $A^{s'}$ are the watershed area above the s and s′ sites. Q_t^s and $Q_t^{s'}$ are the streamflow at sites s and s′ in period t, respectively. Equation (6.10) is not applicable when the change in spatial distribution of rainfall and runoff is significant. This method provides an estimate of the annual average flow at ungauged site s, but seasonal or monthly streamflow can be derived using the ratio of the flow of each period to the annual flow.

When the records of two gauging stations are available, preferably one upstream and one downstream of the ungauged site, streamflow at the ungauged site can be estimated on the basis of the drainage area as follows (Gupta, 1989):

$$\frac{Q_t^s}{A^s} = \frac{Q_t^a}{A^a} + \left(\frac{\overline{Q}_t^a}{A^a} - \frac{\overline{Q}_t^b}{A^b} \right) \left(\frac{A^s - A^a}{A^b - A^a} \right) \tag{6.11}$$

where Q_t^a, Q_t^b is the streamflow at gauged sites a and b in period t; Q_t^s is the streamflow at ungauged site s; and $\overline{Q}_t^a, \overline{Q}_t^b$ is the average of the entire record at gauged sites a and b. The streamflow at an ungauged site can be estimated on the basis of the distance between the sites as:

$$\frac{Q_t^s}{A^s} = \frac{Q_t^a}{A^a} + \left(\frac{\overline{Q}_t^a}{A^a} - \frac{\overline{Q}_t^b}{A^b} \right) \cdot \frac{L^s}{L^b} \tag{6.12}$$

where L^s is the distance between stations s and a, and L^b is the distance between stations b and a. When more than one appropriate gauged site is available, a weighted average at the gauged site can provide an acceptable estimation of flows of ungauged sites. The weights are selected based on relative physical and hydrologic similarities between site s and other gauged sites, such as the relative physical and hydrologic similarities between site s and other gauged sites; for example, the relative distance between the ungauged point and gauged sites can be used in estimation of relative weights.

Example 6.2

The average monthly streamflow at two gauged stations A and B on a river are measured as 4 and 8.4 m^3/s, respectively. The drainage areas at the gauging stations A and B are 48 and 130 km^2, respectively, and the distance between them is 35 km. Consider an ungauged point that is located 10 km downstream of station A and has a drainage area of 75 km^2. The streamflows in September at stations A and B are 3 and 7 m^3/s, respectively. Estimate the streamflow in September at ungauged point.

Solution: For estimating the streamflow in September at the ungauged station, Eq. (6.10) is used and the streamflow at station A is determined as follows:

$$Q^s_{Sep.} = 3 \times \left(\frac{75}{48}\right) = 4.68 \frac{m^3}{s}$$

Using Eq. (6.11), we obtain:

$$\frac{Q^s_{Sep.}}{75} = \frac{4}{48} + \left(\frac{4}{48} - \frac{8.4}{130}\right) \times \frac{75-48}{130-48}$$

$$Q^s_{Sep.} = 6.71 \frac{m^3}{s}$$

and from Eq. (6.12) we have:

$$\frac{Q^s_{Sep.}}{75} = \frac{4}{48} + \left(\frac{4}{48} - \frac{8.4}{130}\right) \times \frac{10}{35}$$

$$Q^s_{Sep.} = 6.65 \frac{m^3}{s}$$

6.7.1.2 Unit Hydrograph As a Tool

Where the drainage basin is not homogenous, the unit hydrograph is recognized as one of the most effective tools in estimating surface runoff. The unit hydrograph is the hydrograph that results from one centimeter of uniform excess rainfall over the entire watershed at a uniform rate and during a certain time period of rainfall. The theory of unit hydrographs is based on four major assumptions:

1. Rainfall excess of equal duration is assumed to produce hydrographs with equivalent time basis regardless of the intensity of the rain.
2. The direct runoff characteristic is dependent on previous precipitation/rainfall and is assumed to be a single event.
3. Direct runoff ordinates are directly proportional to the depth of excess rainfall (principle of linearity).
4. Spatial and temporal distributions of all rainfalls are assumed to be the same.

When the duration of excess rainfall (T) is very small, the unit hydrograph is referred to as an *instantaneous unit hydrograph* (IUH). When the ordinates of a unit hydrograph represent the ratio of time to peak and peak discharge, this kind of unit hydrograph is a *dimensionless hydrograph*. Some theoretical and empirical formulas that relate peak flow and time to peak to watershed characteristics have

been developed and are referred to as *synthetic unit hydrographs*. A regionalized unit hydrograph for ungauged drainage areas can be derived from a gauged watershed with similar characteristics.

Although the unit hydrograph has been used in most rainfall-runoff analyses, care should be taken in applying the unit hydrograph when the principle of linearity is violated. For example, for long-duration rainfalls that vary in intensity, the principle of linearity can be justified. Spatial distribution of rainfall can also change the shape of a hydrograph.

The unit hydrograph can be applied to calculate direct runoff and the streamflow hydrograph. After estimation of abstractions, the excess rainfall hyetograph is calculated from the rainfall hyetograph. The direct runoff, Q_n, can be computed from a given unit hydrograph, U_{n-m+1}, and excess rainfall, P_m, using the following discrete convolution equation:

$$Q_n = \sum_{m=1}^{n} P_m \cdot U_{n-m+1} \tag{6.13}$$

where U_{n-m+1} is the unit hydrograph ordinate. n indicates the time step in which the direct runoff is estimated.

6.7.1.3 Potential Evapotranspiration Estimation: Thornthwaite Method

Different methods have been developed by various researchers in order to estimate the potential evapotraspiration (ET^p), which is a major source of water loss. Thornthwaite and Mather (1955) developed a method in which the input data are precipitation (mm), monthly average temperature (°C), and geographic latitude. In this method, a temporal index (i_m) is estimated for each month (m):

$$\begin{aligned} i_m &= \left(\theta_m / 5\right)^{1.514} && \text{if} \quad \theta_m > 0 \\ i_m &= 0 && \text{otherwise} \end{aligned} \tag{6.14}$$

where θ_m is average air temperature (ºC) in month m. The monthly potential evapotranspiration rate (ET_m^p) is then estimated using the following equation:

$$I = \sum_{m=1}^{12} i_m \tag{6.15}$$

$$a = 6.75 * 10^{-7} I^3 - 7.71 * 10^{-5} I^2 + 1.79 * 10^{-2} I + 0.49239 \tag{6.16}$$

$$ET_m^p = 16(10\theta_m / I)^a f_\lambda \tag{6.17}$$

TABLE 6.2
Average Monthly Temperature (θ) and Temporal Index of Thornthwaite Method

Month	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Temperature (°C)	–4	–2	5.5	12	18	25	28	23	18	10	7	3
i	0	0	1.15	3.75	6.91	11.36	13.48	10.01	6.91	2.85	1.66	0.46

where:

f_λ is a correction factor based on variation of daylight hours in different geographic latitudes.
ET_m^p is the monthly potential evapotranspiration (mm).
I is the annual heat index, excluding subzero temperatures.

Example 6.3

Table 6.2 provides average monthly temperatures in a basin and values of i calculated using Eq. (6.14). Estimate the monthly evapotranspiration in August.

Solution: The annual heat index I is determined as:

$$I = \sum_{m=1}^{12} i_{m,} = 58.5$$

The monthly potential evaporation in August is then estimated assuming $f_\lambda = 1.03$:

$$a = (6.75\times10^{-7})58.5^3 - (7.71\times10^{-5})58.5^2 + (1.792\times10^{-2})58.5 + 0.492 = 1.41$$

$$EP_{Aug.}^P = 16(1.03)\left(\frac{10\times23}{58.5}\right)^{1.41} = 113.6\ mm$$

6.7.1.4 Infiltration Estimation: Green–Ampt Method

Green and Ampt (1911) developed a method for estimation of the infiltration rate that is a conceptual representation of the infiltration process. It is assumed that the wetting front is a sharp boundary dividing the soil with moisture content θ_i below from saturated soil with moisture content η above. Using Darcy's law, the following equation for estimating cumulative infiltration, $F(t)$, may be solved by the successive substitution method:

$$F(t) = Kt + \Psi(\eta - \theta_i)\ln\left(1 + \frac{F(t)}{\Psi(\eta - \theta_i)}\right) \tag{6.18}$$

where K is the hydraulic conductivity (cm/hr), Ψ is the wetting front soil suction head (cm), and η is porosity.

Example 6.4

Consider a silt loam soil with porosity 0.5 and initial moisture content 0.16. Assume that the ponding depth on the surface is negligible, K = 0.65 cm/hr, and Ψ = 16.7 cm. In order to estimate the cumulative infiltration (F) after 2 hr of rainfall, Eq. (6.18) can be used as follows:

$$F(2) = 0.65 \times 2 + 16.7(0.5 - 0.16) \times \ln\left(1 + \frac{F(2)}{16.7(0.5 - 0.16)}\right)$$

$$F(2) = 4.75\ cm$$

$F(2)$ is obtained by the method of successive substitution.

6.7.2 Simulation Models

Simulation is the process of mimicking the dynamic behavior of systems over time. The results of simulation models can be used to describe future possible states of a system. Simulation models have been widely used in the area of water resources planning and management to analyze the impacts of development policies on the performance of systems. These models can be classified into two categories:

1. Hydrologic simulation models
2. Object-oriented programming models

A brief explanation of each category is presented in the following sections, and some examples of simulation models that have been widely used by different investigators are provided.

6.7.2.1 Hydrologic Simulation Models

Hydrologic simulation models have been developed based on the empirical and analytical equations used to model the interrelations and water exchange between surface and subsurface water bodies in the hydrologic cycle; some of these equations were discussed briefly in the previous sections. Simulation models may also incorporate statistical correlations, empirical formulas, and probabilistic estimating methods. In the following sections, some of the most widely used simulation models, including HEC-1, HEC-2, HEC-5, HEC-6, and MIKE-11, are briefly explained.

6.7.2.1.1 HEC-1 Package

The HEC-1, which is a flood hydrograph estimation package, calculates discharge hydrographs from single-storm events for basins of different degrees of complexity. The first version of this model was developed by the Hydrologic Engineering Center of the U.S. Army Corps of Engineers in 1967. The program has several optional capabilities, including parameter estimation and model calibration, and uses the

optimization method, multiplan–multiflood analyses, precipitation depth/area computation, dam-break analyses, and flood damage analyses (Hoggan, 1997). The main capabilities of this model can be summarized as follows:

- Optimal estimation of unit hydrograph, loss rate, and streamflow routing parameters from measured data
- Simulation of watershed runoff and streamflow from historical or design rainfall
- Computation of damage frequency curve and expected annual damages for various locations and multiple flood control plans
- Simulation of reservoir outflow for dam safety analysis

Simulation of the watershed for rainfall-runoff modeling, an important component of river-basin modeling, includes runoff estimation, river and reservoir routing, and combining hydrographs. Dividing the river basin into smaller sub-basins with physical and hydrologic parameters that have lower spatial variability is a primary and basic step for rainfall-runoff modeling. A basin with numerous tributaries and spatial variability in land use and land cover should be broken down into smaller sub-basins to fit the constraints and assumptions of the model. Also, it may be necessary to obtain the information required in terms of level of detail and location as dictated by the objectives of the study. For this purpose, the following items should be considered:

- Historical records of gauging stations at the outlets of sub-basins can significantly improve the accuracy of the calibration process for each sub-basin and the entire watershed. Results of comparisons between observed and computed streamflows can be used in estimating model parameters.
- The model parameters are lumped; therefore, the uniformity of physical and hydrologic characteristics in each sub-basin is an important constraint in defining sub-basins.
- Spatial distributions of rain gauges in different sub-basins should be identified.
- Special precipitation patterns that may affect parts of basin should be determined.

Figure 6.7 shows a schematic of sub-basins and the rainfall-runoff modeling process using the HEC-1 model. The following steps should be considered when using this package for rainfall-runoff modeling:

- Analysis of watershed physiography, including delineation of basin and sub-basin boundaries
- Estimation of basic physical parameters of the sub-basins including:
  - Area
  - Perimeter
  - Slope of the main channel
  - Shape factor

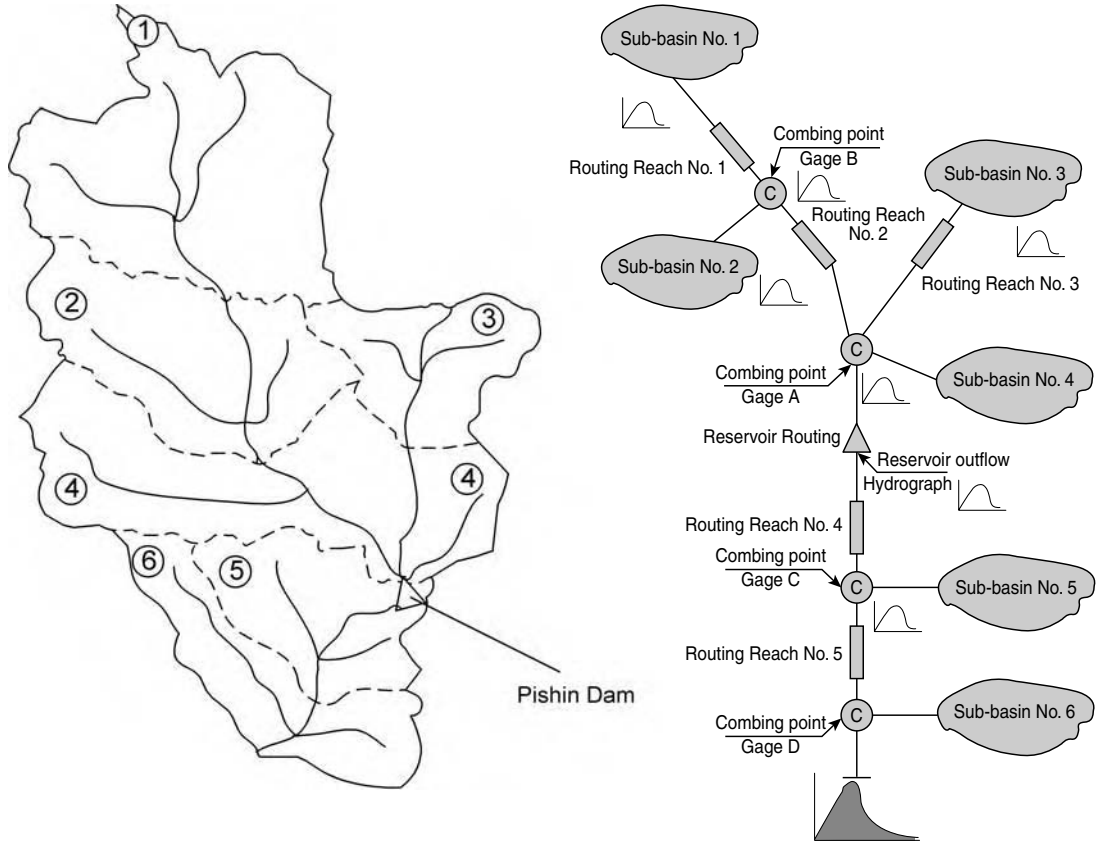


FIGURE 6.7 Schematic of rainfall-runoff model using watershed simulation model.

- Estimation of basic hydrologic parameters of the sub-basins including:
  - Time of concentration (Snyder method or Linsley method)
  - Storage coefficient (Linsley method)
- Snowmelt estimation:
  - Degree-day method
  - Energy balance method
- Interception and infiltration estimation:
  - Initial and constant method (initial loss volume is satisfied, then constant loss rate begins)
  - HEC exponential method (loss function is related to antecedent soil moisture condition and is a continuous function of soil wetness)
  - SCS curve number method
  - Horton's equation method
  - Green–Ampt method
- River and reservoir routing:
  - Maskingum method
  - Maskingum–Cunge method
  - Kinematic wave method
  - Progressive average-lag (Straddle–Stagger) method
  - Modified-plus method

HEC-1 is capable of simulating snowfall and snowmelt. Up to ten elevation zones of equal increments can be specified in each sub-basin. Temperature data at the lowest elevation zone are reduced by the lapse rate (°C or °F per unit change in elevation). The model determines whether precipitation falls as snow or as rain by using the melt temperature (usually ±2°C), which is another input to the model (Bedient and Huber, 1988). A number of methods are available in this package to be used for surface runoff computation:

- Direct input of the unit hydrograph
- Snyder unit hydrograph method
- Clark hydrograph method
- SCS method
- Kinematic wave for overland hydrograph

An automatic optimization procedure is also provided in this package that is a useful tool for trial-and-error parameter selection. Figure 6.8 shows a schematic diagram of parameter estimation in HEC-1. The objective function is to minimize the error index subject to certain constraints (ranges) on the parameters:

$$STDER = \sum_{i=1}^{N} \left(QR_i - QC_i\right)^2 \left(\frac{WT}{N}\right) \tag{6.19}$$

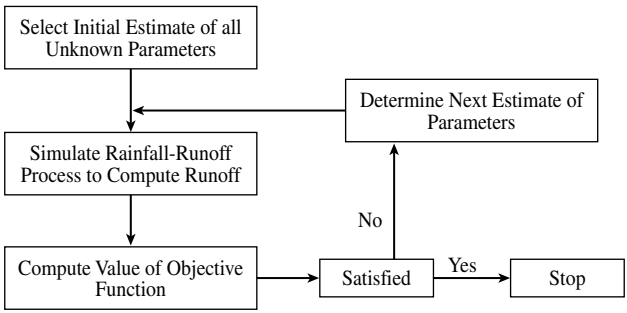


FIGURE 6.8 Schematic of parameter estimation in HEC-1. (From Hoggan, D. H., *Floodplain Hydrology and Hydraulics*, McGraw-Hill, New York, 1997. With permission.)

where:

STDER is the error index.
QR_i is the observed hydrograph ordinate for time i.
QC_i is the computed ordinate for time i from HEC-1.
WT is a weighting function that emphasizes accurate reproduction of peak flows.
N is the total number of hydrograph ordinates.

Frequency analysis and regionalization of hydrologic parameters are the bases of streamflow estimation in ungauged watersheds. The experience and knowledge of the analyst in regard to the model parameters are important when using this model for ungauged watersheds. A more detailed description of this model is presented in the user's manual for this package (Hydrologic Engineering Center, 1981).

6.7.2.1.2 HEC-2

The water surface profile can be computed by the HEC-2 package, which was initially developed in the 1970s by the Hydrologic Engineering Center (1982a) of the U.S. Army Corps of Engineers. The computational procedure is based on solution of one-dimensional energy equations. The following assumptions are considered in development of this model:

- Steady-state flow
- Gradually varied flow
- One-dimensional flow with correction for horizontal velocity distribution
- Small channel slope
- Constant friction slope (averaged) between two adjacent cross sections
- Rigid boundary conditions

Data requirements include flow regime, starting elevation, discharge, loss coefficients, cross-sectional geometry, and reach lengths. An overview of input data is shown in Figure 6.9. In HEC-2, the following two equations are used:

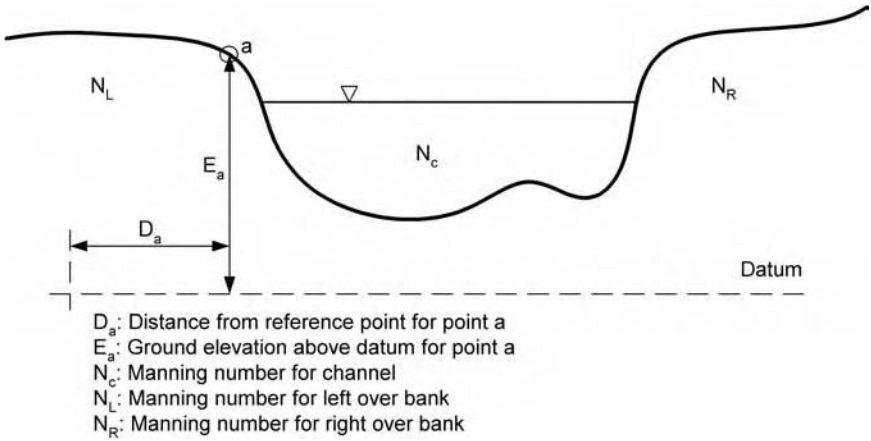


FIGURE 6.9 Overview of HEC-2 input data.

$$WS_2 + \frac{\alpha_2 V_2^2}{2g} = WS_1 + \frac{\alpha_1 V_1^2}{2g} + h_e \tag{6.20}$$

$$h_e = L\bar{S}_f + C\left(\frac{\alpha_2 V_2^2}{2g} - \frac{\alpha_1 V_1^2}{2g}\right) \tag{6.21}$$

where:

WS_1 and WS_2 are the water surface elevations at the ends of the reach.
V_1 and V_2 are the mean velocities (total discharge/total flow area) at the ends of the reach.
α_1 and α_2 are the velocity coefficients for flow at the ends of the reach.
h_e is the energy head loss.
L is the discharge-weighted reach length.
$\bar{S}_f$ is the representative friction slope for the reach.
C is the expansion or contraction loss coefficient.

Four equations in HEC-2 approximate the friction loss between two cross-sections:

$$\bar{S}_f = \left(\frac{Q_1 + Q_2}{K_1 + K_2}\right)^2 \quad \text{average conveyance} \tag{6.22}$$

$$\bar{S}_f = \left(\frac{S_{f1} + S_{f2}}{2}\right) \quad \text{average friction slope} \tag{6.23}$$

$$\bar{S}_f = \sqrt{S_{f1} \cdot S_{f2}} \quad \text{geometric mean friction slope} \tag{6.24}$$

$$\bar{S}_f = \left(\frac{2S_{f_1} \cdot S_{f2}}{S_{f1} + S_{f2}} \right) \tag{6.25}$$

where K_1 and K_2 represent the conveyance at the beginning and end of the reach. Conveyance is defined from Manning's equation as:

$$K = \frac{AR^{2/3}}{n} \tag{6.26}$$

The total conveyance for a cross-section is obtained by summing the conveyance from the left and right overbanks and the channel. The velocity coefficient, α, is estimated using the following relation:

$$\alpha = \left(\frac{A_T^2}{K_T^3} \right) \left(\frac{K_{LOB}^3}{A_{LOB}^2} + \frac{K_{CH}^3}{A_{CH}^2} + \frac{K_{ROB}^3}{A_{ROB}^2} \right) \tag{6.27}$$

where the subscript *T* is the cross-sectional total, *LOB* is the left overbank, *CH* is the channel, and *ROB* is the right overbank. HEC-2 can compute up to 14 profiles in a single run using the same cross-sectional data. Critical depth is also calculated automatically for all cross-sections. For more details about this model, see the HEC-2 user's manual (Hydrologic Engineering Center, 1982).

6.7.2.1.3 HEC-5

HEC-5 was developed in the 1970s at the Hydrologic Engineering Center (1982b) of the U.S. Army Corps of Engineers and is a simulation model for flood control and conservation systems. This program was developed to assist in planning studies for evaluating proposed reservoirs in a system and to assist in estimating the flood control and conservation storage requirements for each project recommended for the system. It is also useful for selecting the proper reservoir releases throughout the system during flood emergencies. The program may be used to determine:

- Flood control and conservation storage requirements for each reservoir in the system
- Influence of a system of reservoirs on the spatial and temporal distribution of runoff in a basin
- Evaluation of operational criteria for both flood control and conservation (including hydropower) for a system of reservoirs
- Expected annual flood damages, system costs, and system net benefits for flood damage reduction

- Design of a system of existing and proposed reservoirs or other alternatives (including nonstructural alternatives) that results in the maximum net flood control benefits, as determined by simulation runs for selected alternative systems

The following methods are available in this package for river and reservoir routing:

- Stream routing:
  - Straddle–Stragger
  - Tatum
  - Muskingum
  - Modified plus
  - Working R&D
- Reservoir routing:
  - Accounting method
  - Surcharge routing
  - Emergency release

Expected annual damages can be computed for different control points using one or more ratios for each of several historical or synthetic floods. Expected annual damages, costs, and net benefits for non-reservoir alternatives such as levees, channel improvements, and nonstructural alternatives, including floodproofing and floodplain zoning, can also be evaluated.

In addition to flood control operation, conservation operation may also be evaluated by this model to provide minimum flows at downstream locations based on required reservoir storage. Hydropower reservoirs can also be evaluated by HEC-5 in order to meet on-site monthly, weekly, daily, and hourly energy requirements or allocated system energy. For this purpose, the tailwater rating curve and power generation efficiency are inputs to the model. Hydropower reservoirs can also be operated based on a rule curve relating plant factors to a percent of conservation storage. Both constant and varied diversions are included as a function of inflows or reservoir storage. Benefits for hydropower reservoirs can be calculated based on input rates for energy, secondary energy, and power shortages (negative benefits). Table 6.3 shows the reservoir operation priorities in the HEC-5 model (Hydrologic Engineering Center, 1982), and Figure 6.10 provides various reservoir storage levels in the model.

6.7.2.1.4 HEC-6

HEC-6, which was first developed in the 1970s at the Hydrologic Engineering Center of the U.S. Army Corps of Engineers, is a one-dimensional, movable-boundary, open-channel-flow, numerical model. This model is designed to simulate and predict changes in river profiles resulting from scour and/or deposition over moderate time periods. A continuous flow record is partitioned into a series of steady flows of variable discharge and duration. For each flow, a water surface profile is calculated,

TABLE 6.3
Reservoir Operation Priority in HEC-5 Model

Condition	Normal Priority	Optional Priority
Flooding at downstream location	No release for power requirements	Release for primary power
When primary power releases can be made without increasing flooding downstream	Release down to top of buffer pool	Release down to top of inactive pool
Flooding at downstream location	No release for minimum flow	Release minimum desired flow
If minimum *desired* flows can be made without increasing flooding downstream	Release minimum flow between top of conservation and top of buffer pool	Same as normal
If minimum *required* flows can be made without increasing flooding downstream	Release minimum flow between top of conservation and top of inactive pool	Same as normal
Diversions from reservoirs (expect when diversion is a function of storage)	Divert down to top of buffer pool	Divert down to top of inactive pool

Source: Hydrologic Engineering Center, *HEC-5 Simulation of Flood Control and Conservation Systems*, U.S. Army Corps of Engineers, Davis, CA, 1982.

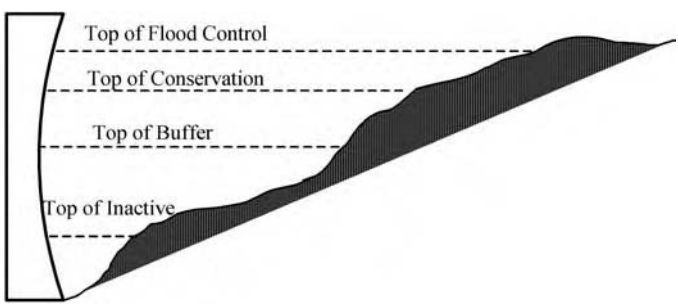


FIGURE 6.10 Reservoir storage levels in HEC-5 model.

thereby providing energy slope, velocity, depth, etc. at each cross-section. Potential sediment transport rates are then computed at each section. These rates, combined with the duration of the flow, permit a volumetric accounting of sediment within each reach. The amount of scour or deposition at each section is then computed and the cross-sections are adjusted accordingly. The computations then proceed to the next flow in the sequence, and the cycle is repeated beginning with the updated geometry.

Separation of sediment deposition from the hydraulics of flow is valid in some circumstances; for example, deposition in deep reservoirs can usually be characterized as a progressive reduction in storage capacity if the material is rarely entrained once

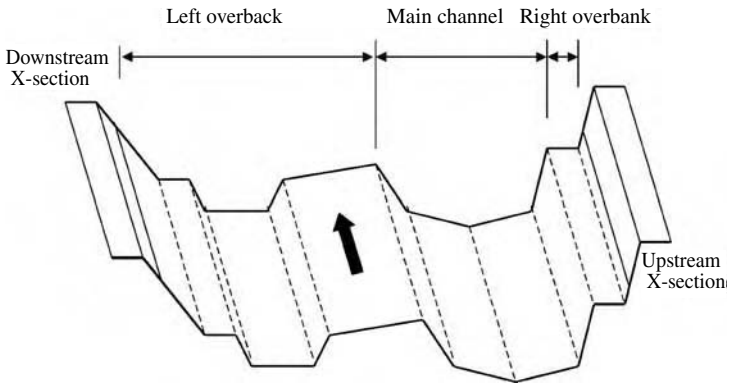


FIGURE 6.11 Schematic illustrating the main channel and right and left overbank reaches between consecutive cross-section in HEC-6. (From Hydrologic Engineering Center, *HEC-6 Scour and Deposition in Rivers and Reservoirs*, U.S. Army Corps of Engineers, Davis, CA, 1993.)

it is deposited. HEC-6 is capable of predicting sediment behavior in shallow reservoirs and most rivers, however, which requires that the interactions between the flow hydraulics, sediment transport, channel roughness, and related changes in boundary geometry be considered.

The procedure for clay and silt deposition is based on settling velocity. Different methods are available for sediment transport functioning of bed material load.

A river system that is modeled by HEC-6 consists of main stem, tributaries, and local inflow/outflow. A one-dimensional energy equation is used for water surface profile computations. Manning's equation and n values for overbank and channel areas may be specified by discharge or elevation (Figure 6.11). Expansion and contraction losses are included in the determination of energy losses. The energy loss coefficients may be changed at any cross-section.

For each discharge in a hydrograph, the downstream water surface elevation can be determined by either a user-specified rating curve or a time-dependent water surface elevation. Internal boundary conditions can be imposed on the solution. The downstream rating curve can be changed at any time, as can internal boundary conditions.

Flow conveyance limits, containment of flow by levees, ineffective flow areas, and overtopping of levees are simulated in a manner similar to HEC-2. Supercritical flow is approximated by normal depth; therefore, sediment transport phenomena occurring in supercritical reaches are simplified in HEC-6. For more details about this package, see the user's manual (Hydrologic Engineering Center, 1993).

6.7.2.1.5 MIKE-11

MIKE-11 is a software package developed at the Danish Hydraulic Institute (1992) for the simulation of flows, sediment transport, and water quality in estuaries, rivers, irrigation systems, and similar water bodies. It consists of four major parts:

1. Rainfall-runoff process modeling
2. Modeling of river and estuary hydraulics
3. Noncohesive sediment processes
4. Quality processes and advection-dispersion modeling

Figure 6.12 shows a typical MIKE-11 simulation. The rainfall-runoff process is modeled using the NAM model, which was developed by the Institute of the Hydrodynamics and Hydraulic Engineering at the Technical University of Denmark. The NAM model considers evapotranspiration, groundwater recharge, base flow, and groundwater storage and interflows. It also can be used to describe the effects of irrigation within a watershed, including:

- The overall water balance of the watershed, which is mainly affected by increased evapotranspiration and possible external water resources for irrigation
- Local infiltration and groundwater recharge in irrigated areas
- Distribution of watershed runoff on different components (overland flow, interflow, baseflow), which may be influenced by increased infiltration in irrigated areas

Water transfer in the root zone is also considered when modeling evapotranspiration. The meteorological data used in the NAM model are rainfall, potential evapotranspiration, and temperature.

In the hydraulic modeling module, MIKE-11 uses an implicit, finite-difference model for computation of unsteady flows in rivers and estuaries. The model is also capable of describing subcritical and supercritical flow conditions by means of a numerical scheme that adapts to local flow conditions. Computational modules are included for description of flow over hydraulic structures. The formulations can be applied to loop networks and quasi two-dimensional flow simulation on floodplains. The computational scheme is applicable for vertically homogeneous flow conditions ranging from steep river flows to tide-influenced estuaries.

The noncohesive sediment transport module can be run in two modes: explicit and morphological. In the explicit mode, output is required from the hydrodynamic module in terms of discharge water levels both in time and space. Results are provided in the form of volume transport rates and accumulated volumes of deposition or erosion. The explicit mode is useful where significant morphological changes are unlikely to occur. In the morphological mode, sediment transport is calculated in tandem with the hydrodynamic module. The feedback between these modules is achieved through solution of the sediment continuity equation and updating the bed resistance and subsequent sediment transport. Results of this mode are in the form of sediment transport rates, bed-level changes, resistance number, and dune dimensions.

The advection–dispersion modeling module is based on the one-dimensional equation of conservation of mass of a dissolved or suspended material. The advection–dispersion equation is solved numerically using an implicit finite difference scheme which, in principle, is unconditionally stable and has negligible numerical dispersion. A correction term has been introduced in order to reduce the third-order truncation error. This correction term makes it possible to simulate dispersion/convection of

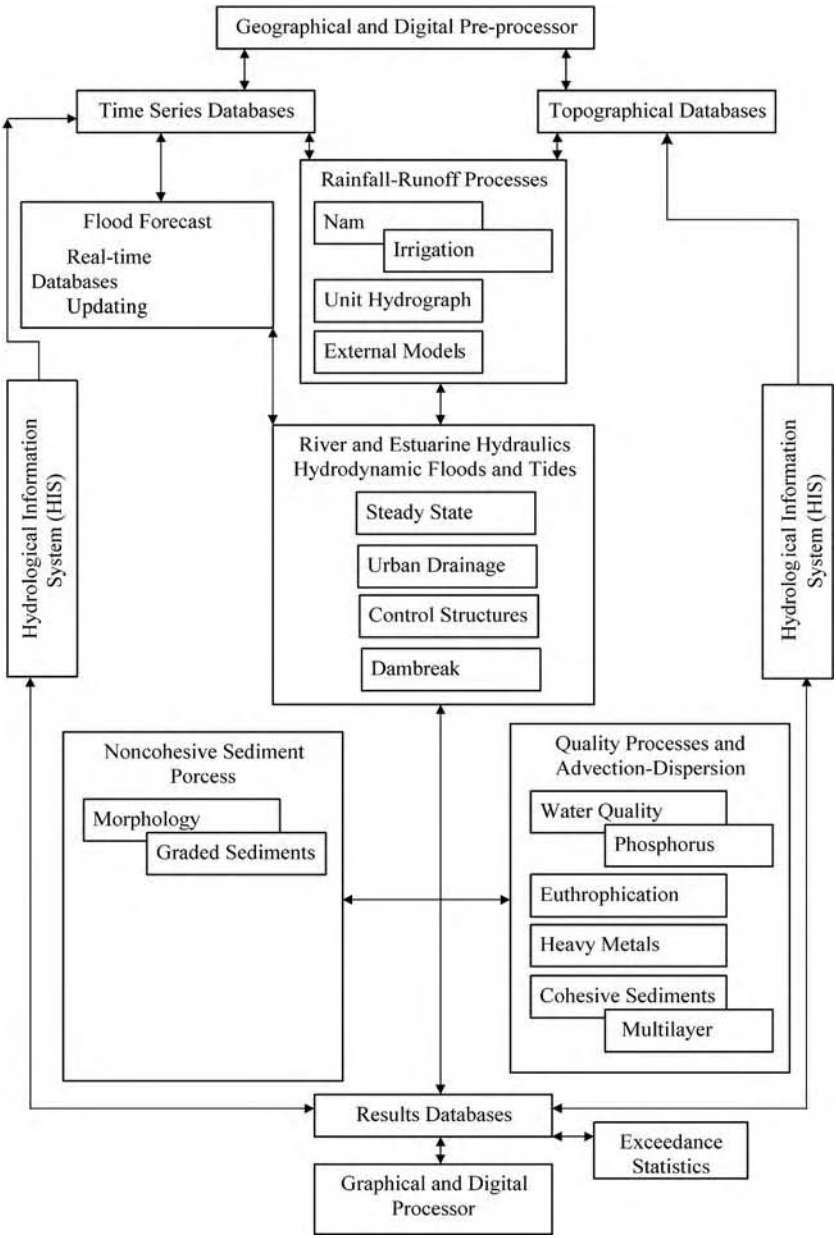


FIGURE 6.12 Typical MIKE-11 simulation. (From Danish Hydraulic Institute, *MIKE-11: A Microcomputer Based Modeling System for Rivers and Channels — Reference Manual*, Denmark, 1992.)

concentration profiles with very steep fronts. The water quality module, which is coupled with the advection–dispersion module, deals with basic aspects of river water quality as influenced by human activities, such as oxygen depletion and ammonia levels resulting from organic matter loadings.

For more details about the models used in the various MIKE-11 modules and their limitations, see the reference manual (Danish Hydraulic Institute, 1992).

6.7.2.1.6 IRAS

IRAS (Interactive River-Aquifer Simulation Program) is a generic model developed by Cornell University in 1995. This program can simulate water flows, water storage volumes, water qualities, and hydroelectric power and energy produced and consumed throughout a surface water or an independent surface and groundwater system over time. IRAS has been developed to assist those interested in evaluating the performance or impact of alternative design and operation policies of regional water resources systems.

The system to be simulated by this model should be represented by a network of connected nodes and links. The nodes of the network can represent aquifers, gauge sites, natural lakes, reservoirs, wetlands, confluences, and diversions. A single node may be a combination of any of these nodes. The links of the network can be:

- Unidirectional (flows only in one direction)
- Bidirectional (flows in either direction or flows dependent on changing surface elevation or pressure head differences)

IRAS can simulate up to 10 separate independent water quality constituents, and the system to be simulated can include up to 60 stream and river reaches and up to 60 sites that can represent any feasible combination of aquifers, gauge and monitoring sites, consumption sites, natural lakes, wetlands, reservoirs, confluences, diversion and withdrawal sites, hydropower and pumping sites, and wastewater effluent discharge sites. The time resolution can be defined by the user and can include within-year periods and their durations, as well as the number of simulation time steps contained in each within-year period. If the demand for water at various sites depends on local rainfall, streamflow, or storage at those sites, demand-driven targets can be defined and water allocations to satisfy these demands can vary from year to year.

The reservoir release policy based on storage volume zones (supply-driven policies) should be defined by the user. For this purpose, storage volume boundaries and releases associated with the storage volumes in each of the storage volume zones should be defined by the user. Figure 6.13 shows a reservoir release rule based on storage volumes in the program. The program is also capable of simulating reservoir operation based on demand deficits. Releases based on demand deficits, prorated over the remainder of the within-year period and then multiplied by a source multiplier for the reservoir, are added to the release based on supply-driven policies.

6.7.2.2 Object-Oriented Simulation Models

Applying object-oriented modeling when utilizing a system dynamic (SD) approach for modeling water resources systems is a new way of thinking about modeling real-

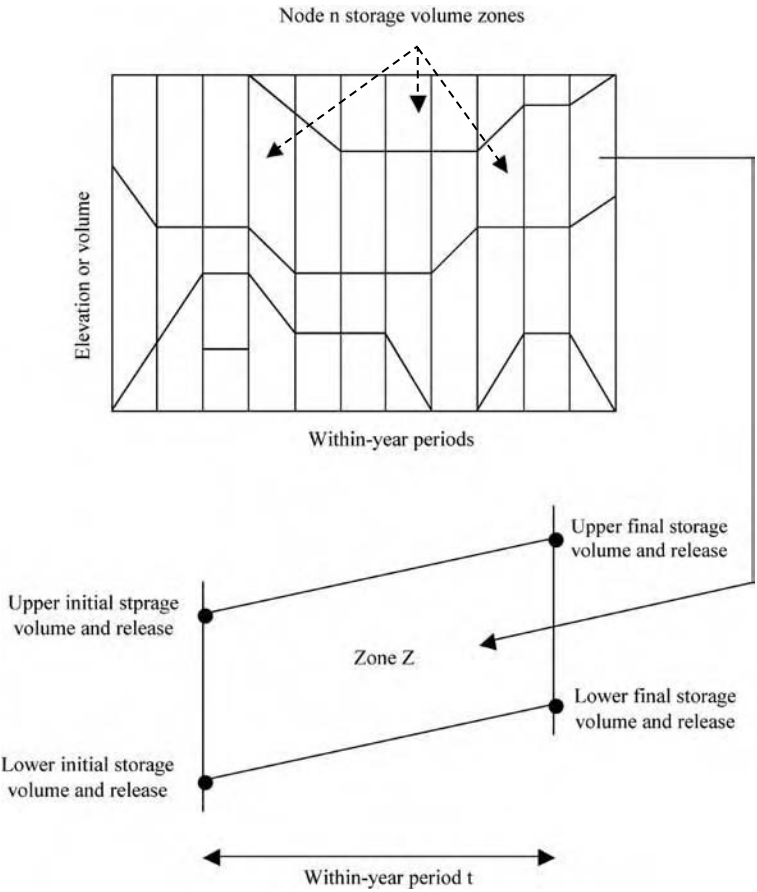


FIGURE 6.13 Reservoir release rule based on storage volumes in IRAS. (From Loucks, D. P. et al., *IRAS: Interactive River-Aquifer Simulation — Program Description and Operation*, Resources Planning Associate, Inc., Ithaca, NY, 1995. With permission.)

world concepts. SD models provide insight into feedback processes and therefore give system users a better understanding of the dynamic behavior of systems. The SD approach is based on the theory of feedback processes. In a feedback system, future behavior is influenced by its past. Feedback systems can be classified into two different classes (Ahmad and Simonovic, 2000):

- *Negative feedback*, which seeks a goal and responds as a consequence of failing to achieve the goal
- *Positive feedback*, which generates growth processes where action builds a result that generates still greater action

The feedback loop indicates how a system might behave because of its internal feedback loops and the effects that positive and negative feedback loops have on a system. By separating policy questions from data, object-oriented modeling and a

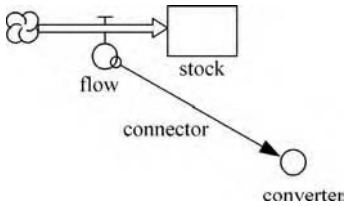


FIGURE 6.14 Basic building blocks of an object-oriented simulation model. (From Simonovic, S. P., *Water Int.*, 25(1), 76–88, 2000. With permission.)

system dynamic approach make the model results functionally transparent to all parties involved in the water management program. The proposed approach is flexible and transparent and allows for easy involvement of stakeholders in the process of decision analysis about water resources (Simonovic, 2000).

Figure 6.14 shows the basic elements of an object-oriented simulation model. *Stock* represents accumulations and can be used as a source — for example, water stored in the reservoir. The accumulation needs flow (inflow to the reservoir) is modeled by *Flow*. Flow and Stock are inseparable and comprise a minimum set of elements necessary to describe dynamics. *Converters* convert input to outputs; they can represent information or material quantities. *Connectors* link stocks to converters, stocks to flow regulators, and converters to other converters.

Computer software tools such as STELLA, DYNAMO, VENSIM, and POWERSIM can be used for the development of object-oriented simulation models for complex water resources systems. For more details see High Performance Systems (1992), Lyneis et al. (1994), Ventana Systems (1996), and Powersim Corp. (1996).

6.7.3 Optimization Models

Optimization models are used to find the best way to meet different objectives of river system management. Such models might be used, for example, to determine the volume of water that should be allocated to various demand points along the river or to select the treatment efficiency that should be applied to effluents discharging to the river in order to meet environmental constraints. Some examples of the application of optimization models in river system management are discussed in the following sections.

6.7.3.1 Water Allocation in Multi-User Systems

The costs of excluding potential (excluding cost) beneficiaries from exploiting water resources along the river and saving it for supplying water rights are an important component of total river management costs. Exclusion costs are highly correlated with the level of relative scarcity; where water is scarce, the costs of monitoring and enforcing regulations for water allocation among users are higher than where water is plentiful relative to demand. Optimization models are useful tools for development of water allocation heuristics when water is scarce. The objective function of water allocation problems can be the minimization of total costs of operation of systems as follows:

$$\text{Min} \quad \sum_{t=1}^{n}\sum_{i=1}^{ND}\left(C_{t,i} - B_{t,i}\right) \tag{6.28}$$

where ND is the number of demand points and $C_{t,i}$ is the total cost of water allocation to user i in time t, which can be estimated as the costs of water diversion ($C_{i,t}^{div}$) plus exclusion costs ($C_{i,t}^{exc}$) associated with water allocation and controlling the water consumption by user i:

$$C_{t,i} = C_{t,i}^{div} + C_{t,i}^{exc} \tag{6.29}$$

where $B_{t,i}$ is the total benefit of water allocation to user i in time t, which can be determined based on the price of water allocated to various users. As mentioned in the previous sections, deterministic and stochastic models can be used for optimization of river–reservoir systems.

As an example of stochastic modeling, consider a water management authority that should deliver water from a river to municipal and industrial water users. Let D^1 and D^2 be the water demands of the municipal and the industrial water users, respectively. If the demands are supplied, the resulting benefits to the water management authority can be estimated considering the price of water for these uses (P_1, P_2). However, if the water rights are not supplied, water must be delivered from an alternative source, which would incur greater costs, or the costs associated with rationing must be incorporated (C_1, C_2). Let Q be the random variable representing the available water. The objective function of stochastic optimization model can be formulated as follows:

$$\text{Maximize} \quad \left(\sum_{i=1}^{2} P_i D^i\right) - E\left(\sum_{i=1}^{2} C_i F_{i,Q}\right) \tag{6.30}$$

where $F_{i,Q}$ is the amount by which demand D^i is not met when the river flow is Q. To solve this problem with dynamic programming or linear programming techniques, the probability distribution of Q should be approximated by a discrete distribution. If Q take values q_j with probability p_j ($j = 1, \ldots, n$), the objective function can be reformulated as follows (Loucks et al., 1981):

$$\text{Maximize} \quad \left(\sum_{i=1}^{2} P_i D^i\right) - \left(\sum_{i=1}^{2}\sum_{j=1}^{n} p_j C_i F_{i,j}\right) \tag{6.31}$$

$$\text{Subject to} \quad q_j \geq D_1 + D_2 - F_{1,j} - F_{2,j} \tag{6.32}$$

$$D_i \geq F_{i,j} \tag{6.33}$$

The above constraints should be repeated for all values of j.

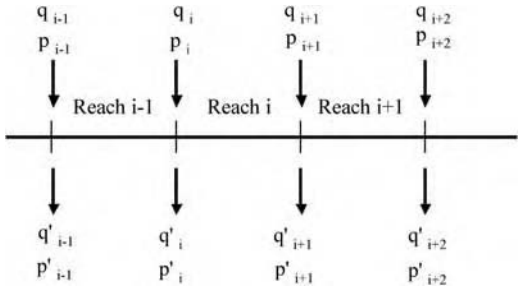


FIGURE 6.15 Schematic of multireach river.

6.7.3.2 River Quality Management

Optimization models have been widely used for development of cost-effective water pollution control strategies. These models usually focus on different measures of quality of water in receiving wastewaters. Such models are particularly useful for management efforts that include the entire watershed because they are capable of accounting for many interactions that occur among the various sources of pollution in a watershed, the effects of pollution on water quality, and the costs of options for reducing pollution discharges (ReVelle and McGarity, 1997).

Consider a multireach, multidischarge river as shown in Figure 6.15, where q_i and p_i are quantity and quality of water discharged to the river at the beginning of reach i, respectively. q_i' and p_i' are the same indicators for water diverted at the beginning of reach i. The optimization model for finding the best treatment efficiency for each of the discharges (e_i) can be formulated as follows:

$$\text{Min} \quad \sum_{i=1}^{n} C_i(e_i) \tag{6.34}$$

$$\text{Subject to:} \quad Q_i = Q_{i-1} + q_i - q_i' \tag{6.35}$$

$$P_i Q_i = Q_{i-1}(P_{i-1}A_{i-1}) + e_i q_i p_i - q_i' p_i' \tag{6.36}$$

$$P_i \le P_{\max} \tag{6.37}$$

where:

C_i is the cost of applying treatment with a specific efficiency on discharge i.
Q_i is the river discharge in reach i.
P_i is the river water quality in reach i.
A_i is the percent of decrease in the quality indicator (P_i) due to self-purification of the river in reach i.
P_{max} is the maximum or minimum allowed value of the quality indicator.

Details of water quality management models are presented in Chapter 9.

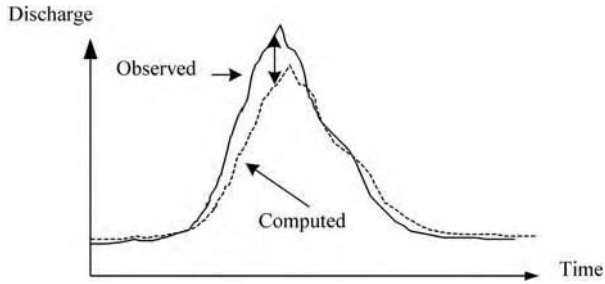


FIGURE 6.16 Comparison between computed and observed hydrographs for calibration of rainfall-runoff models.

6.7.3.3 Calibration of Hydrologic Simulation Models

Model calibration refers to adjusting the model and parameters to bring the model outputs as close to the observed values as possible; however, the parameters of hydrologic models almost always depend on climatic, geographic, and hydrologic characteristics. Because of the large number of factors affecting hydrologic processes and a lack of knowledge about the details of particular hydrologic processes, parameter estimation is a major source of error in modeling. Optimization algorithms can play an important role in improving the calibration process. Application of the optimization algorithm in the HEC-1 package for modeling the rainfall-runoff process was discussed previously; this algorithm is based on minimizing differences between the computed and observed runoff hydrographs, as shown in Figure 6.16.

6.8 CONFLICT ISSUES IN RIVER SYSTEMS

River system planning and management is usually a multiobjective problem, with many objectives in conflict. For example:

- Several demand points are usually located along the river, and water that is supplied to one of these demand points cannot be used by others; therefore, the major conflict issue in river system planning arises when the river flow is less than instream and offstream water requirements.
- Water supply to some demand points, such as agricultural zones, will increase low-quality return flows and endanger aquatic life.
- The treatment efficiency of pollutant dischargers from various sectors (e.g., domestic, industrial, agricultural) can result in conflicts among these sectors stemming from a desire to reduce violations of water quality standards.

River system planning and management conflicts arise when the water demands of different sectors are supplied from one river system. In order to examine this problem, consider a river system that supplies the following demands:

- Domestic water demand
- Agricultural water demand
- Industrial water demand
- In-stream water demand (environmental and recreation water demand)

Water allocation schemes are typically defined and imposed by:

- Department of Water Supply
- Department of Agriculture
- Department of Environmental Protection
- Department of Industries

The priorities and favorable ranges of water supplies for each of these agencies should be considered when formulating conflict resolution problems; therefore, the first step in conflict resolution studies is to recognize all of the conflict issues and the responsible agencies. The next step is to define the acceptable range of allocated water to each sector and its corresponding utility function. The utility function assigns a number between 0 and 1 to each value of allocated water; 0 indicates that this value of allocated water is not acceptable, and 1 indicates that the allocated water completely satisfies the water demands of the sector. The utility functions are developed by decision-makers from the various sectors. The final step is to formulate the conflict resolution problem. For this purpose, different methods can be used, as briefly explained in Chapter 2. A typical conflict resolution problem in river systems is explained in the following example.

Example 6.5

Determine the monthly water allocation for domestic, industrial, agricultural, and recreation demands in the river system shown in Figure 6.17. The average monthly river discharges upstream of the system, within a 2-year time horizon, are presented in Table 6.4. The return flow of domestic and industrial sectors is assumed to be 20% of the allocated water, and the initial volume of the lake is 30 million m^3. The utility functions of the sectors are presented in Table 6.4 and Figure 6.18. The utility values have been normalized between 0 and 1, and higher utility values reflect higher priorities of the decision-makers or sectors. The shape of the graphed utilities is considered to be trapezoidal, and the array (a, b, c, d) in Table 6.4 shows the values of water allocated to the agricultural sector, corresponding to utilities of 0, 1, 1, and 0, respectively.

Solution: The Nash solution of this problem is the unique optimal solution of the following problem:

$$\text{Maximize} \quad \prod_{i=1}^{4} \prod_{t=1}^{12} (f_{i,t} - d_{i,t})^{w_i} \tag{6.38}$$

$$\text{Subject to} \quad d_{i,t} \le f_{i,t} \le f_{i,t}^{*} \qquad \forall i,t \tag{6.39}$$

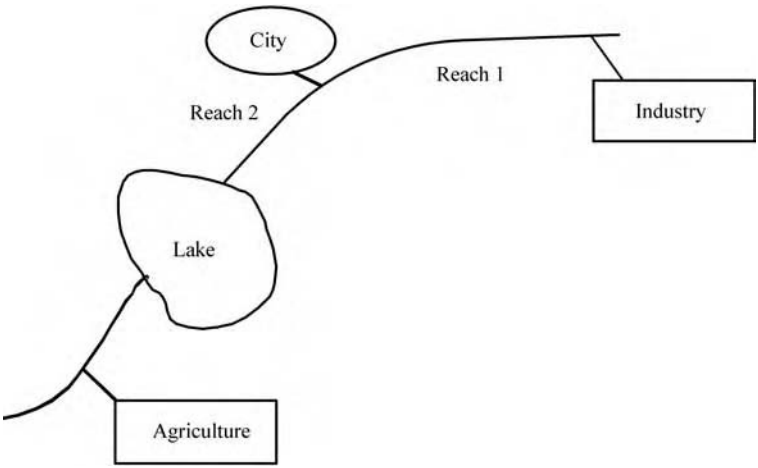


FIGURE 6.17 Components of the river system (for Example 6.5).

TABLE 6.4
Monthly River Discharge and Utility Functions of the Agricultural Sector

Month	Agricultural Demand (million m3/month)	River Discharge Upstream of the System (million m3)
1	(0,10,15,25)	55
2	(0,8,12,20)	53
3	(0,0.01,25,25.01)	55
4	(0,0.01,25,25.01)	60
5	(0,0.01,25,25.01)	65
6	(0,5,10,15)	80
7	(5,10,25,50)	80
8	(15,30,60,90)	85
9	(20,35,60,90)	80
10	(25,35,60,90)	75
11	(30,40,60,90)	73
12	(25,35,60,10)	70

where W_i is the relative weight of sector i, which shows the relative authority of this sector in water allocation; $f_{i,t}$, $f_{i,t}^*$, and $d_{i,t}$ are the utility function, the ideal point, and the disagreement point of sector i in period t, respectively. The relative weights of agencies that are related to agricultural, domestic, industrial, and recreational demands are assumed to be 1. The water allocated to these sectors based on the conflict resolution model is presented in Table 6.5.

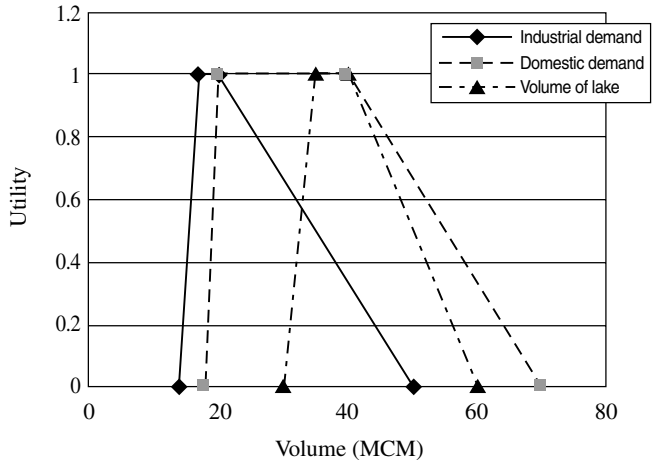


FIGURE 6.18 Utility function of different sectors for water allocated to industry (for Example 6.5).

6.9 PROBLEMS

6.1 The length of the main stream of a watershed from the outlet to the divide is 20 km. The watershed shape parameter (L_c) is 12 km, and the watershed storage coefficient (C_t) is 2.1. Find the lag time of a watershed by Snyder's method.

6.2 Solve the previous problem using Linsley's method. Assume that the slope of the watershed is 10%, and the watershed storage coefficient (C_l) is 1.2.

6.3 Find the storage coefficient for the basin in problem 6.2 with respect to the empirical coefficient, C_r. Assume that the basin area is 50 km^2.

6.4 Results of frequency analysis of historical data for the annual peak flows at the outlet of three basins are as follows:

Return Period (yr)	A		B		C	
	Discharge (m^3/s)	Damage (1000$)	Discharge (m^3/s)	Damage (1000$)	Discharge (m^3/s)	Damage (1000$)
10	240	350	560	250	110	650
25	350	400	650	320	190	680
100	420	620	800	450	250	930
Carrying capacity (m^3/s)	100		400		90	

Find the annual expected damage for all basins.

6.5 In problem 6.4, reducing the flood damage by channel modification has been studied for three basins. Results of this study have shown that the capital cost of channel improvement in each basin can be estimated as a function of flood peak discharge and the river carrying capacity:

$$C_i = \alpha_i \times (Q_i^P - Q_i^C)^3$$

TABLE 6.5
Water Allocation to Different Sectors Based on the Conflict Resolution Model

Month (year 1)	Industrial (million m^3)	Agriculture (million m^3)	Domestic (million m^3)	Lake (million m^3)	Month (year 2)	Industrial (million m^3)	Agriculture (million m^3)	Domestic (million m^3)	Lake (million m^3)
1	17.56	24.78	34.91	13.70	1	26.65	38.15	38.75	8.91
2	17.56	28.13	38.75	10.96	2	18.72	39.82	45.48	9.05
3	23.34	24.78	36.83	8.11	3	19.87	39.82	38.75	11.31
4	18.72	26.45	38.75	19.30	4	33.73	33.14	34.91	6.51
5	19.87	39.82	35.87	14.50	5	18.72	21.44	37.79	12.11
6	24.49	36.48	34.91	5.36	6	19.87	31.47	39.71	9.17
7	18.72	31.47	34.91	22.44	7	17.56	28.13	39.71	19.55
8	17.56	31.47	36.83	39.23	8	19.87	23.11	39.71	44.05
9	23.34	23.11	46.44	35.86	9	23.34	23.11	36.83	31.36
10	17.56	23.11	36.83	35.55	10	19.87	23.11	35.87	31.37
11	21.03	24.78	36.83	35.89	11	16.41	21.44	34.91	30.10
12	18.72	24.78	39.71	31.37	12	16.41	21.44	37.79	29.28

where:

C_i is the cost of channel improvement in basin i (\$).
Q_i^P is the peak discharge in basin i (m^3/sec).
Q_i^C is the carrying capacity of the main river in basin i (m^3/sec).
α_i is the estimated coefficient for basin i.
α is estimated to be 0.001, 0.0005, and 0.01 for basins A, B, and C, respectively. In which basin is the channel improvement more cost effective? The annual maintenance costs are estimated to be equal to 5% of the capital investment required for each of these basins.

6.6 Consider a basin covered by silt loam soil with porosity 0.5 and initial moisture content 0.16. Assume that the hydraulic conductivity is 0.65 cm/hr, and the wetting front soil suction head is 16.7 cm. The area of the basin is 220 km^2. The frequency analysis of rainfall data in this basin is shown in the following table:

Return Period (yr)	Duration (hr)	Intensity (cm/hr)
5	5	1
10	4	2
100	5	5

The flood damages at the outlet of the basin are as follows:

Flood Volume (million m^3)	Flood Damage (10^6 \$)
10	25
20	40
30	55
40	100

Construction of a dam may reduce the flood damages. The annual costs of the dam considering the capital investment and annual costs of operation and maintenance are shown in the following table:

Reservoir volume (million m^3)	Annual Cost (10^6 \$)
5	2.5
10	3
15	4
20	7
40	12

If the evaporation rate is estimated to be 10 mm, find the optimal volume of the reservoir.

6.7 A river supplies water to an industrial complex and agricultural lands located downstream of the complex. The average monthly industrial and irrigation demands and the monthly river flows in a dry year are as follows (numbers are in million m^3):

	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Industrial	8	8.5	8.5	9	9	9.5	9.5	10	9.5	9	8.5	8.5
Irrigation	0	0	20	50	55	70	70	65	20	10	0	0
Monthly flows	25	27	40	48	60	42	25	17	14	18	20	25

The price of water for industrial and irrigation uses is \$90,000 and \$10,000/million m^3, respectively.

(a) Formulate the problem for optimizing the water allocation for this river.

(b) Solve the problem using linear programming.

6.8 In problem 6.7, the shortages should be supplied from groundwater resources. The cost of supplying demand from groundwater is a function of volume of water, which should be extracted. The cost of water extraction for industrial demands is estimated as $C_{ind} = 1000 \cdot x_{ind}^2$, where x_{ind} is the volume of water extraction for industrial complex. The cost of water extraction for irrigation purposes is estimated as $C_{irr} = 250 \cdot x_{irr}^2$, where x_{irr} is the volume of water extraction for irrigation. Formulate the problem for finding the optimal monthly volumes of water allocation.

6.9 In problem 6.7, the industrial complex discharges its wastewater to the river upstream of the agricultural lands. The monthly wastewater discharge rate is about 20% of the water use of the complex in each month. In order to keep the quality of the river flow in an acceptable range for irrigation, the pollution load of the industrial wastewater should be reduced by primary treatment, which has additional cost equal to \$12,000/million m^3.

(a) Formulate the problem for optimizing the water allocation from this river.

(b) Solve the problem using linear programming.

References

Ahmad, S. and Simonovic, S. P., System dynamics modeling of reservoir operations for flood management, *J. Comput. Civil Eng.*, 14(3), 190–198, 2000.

Barrett, E. C. and Curtis, L. F., *Introduction to Environmental Remote Sensing*, Chapman & Hall, London, 1992.

Bedient, Ph. B. and Huber, W. C., *Hydrology and Floodplain Analysis*, Addison-Wesley, Reading, MA, 1988.

Danish Hydraulic Institute, *MIKE-11: A Microcomputer Based Modeling System for Rivers and Channels — Reference Manual*, Denmark, 1992.

Green, W. H. and Ampt, G. A., Studies of soil physics, 1: The flow of air and water through soils, *J. Ag. Sci.*, 4(1), 1–24, 1911.

Gupta, R. S., *Hydrology and Hydraulic Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

High Performance Systems, *Stella II: An Introduction to Systems Thinking*, High Performance Systems, Inc., Hanover, NH, 1992.

Hoggan, D. H., *Floodplain Hydrology and Hydraulics*, McGraw-Hill, New York, 1997.

Hydrologic Engineering Center, *HEC-1 Flood Hydrograph Package: User's Manual and Programmer's Manual* (updated 1985), U.S. Army Corps of Engineers, Davis, CA. 1981.

Hydrologic Engineering Center, *HEC-2 Water Surface Profiles Flood Hydrograph Package: User's Manual*, U.S. Army Corps of Engineers, Davis, CA, 1982a.

Hydrologic Engineering Center, *HEC-5 Simulation of Flood Control and Conservation Systems*, U.S. Army Corps of Engineers, Davis, CA, 1982b.

Hydrologic Engineering Center, *HEC-6 Scour and Deposition in Rivers and Reservoirs*, U.S. Army Corps of Engineers, Davis, CA, 1993.

Linsley, R. K., Kohler, M. A., and Paulhus, J. L. H., *Hydrology for Engineers*, 3rd ed., McGraw-Hill, New York, 1982.

Loucks, D. P., Stedinger, J. R., and Haith, D. A., *Water Resources Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1981.

Loucks, D. P., French, P. N., and Taylor, M. R., *IRAS: Interactive River-Aquifer Simulation — Program Description and Operation*, Resources Planning Associate, Inc., Ithaca, NY, 1995.

Lyneis, J., Kimberly, R., and Todd, S., Professional dynamo: simulation software to facilitate management learning and decision making, in *Modeling for Learning Organizations*, Morecroft, J. and Sterman, J., Eds., Pegasus Computations, Waltham, MA, 1994.

Mays, L. W., *Water Resources Handbook*, McGraw-Hill, New York, 1996.

McCuen, R. H., Hydrologic analysis and design, Prentice-Hall, Englewood Cliffs, NJ, 1998.

Powersim Corp. (1996), *Powersim 2.5 Reference Manual*, Powersim Corporation, Inc., Herndon, VA, 1996.

ReVelle, C. and McGarity, A. E., *Design and Operation of Civil and Environmental Engineering Systems*, John Wiley & Sons, New York, 1997.

Sanders, T. G., Ward, R. C., Loftis, J. C., Steele, T. D., Ardin, D. D., and Yevjevich, V., *Design of Networks for Monitoring Water Quality*, Water Resources Publications, Littleton, CO, 1983.

Simonovic, S. P., Tools for water management: one view of the future, *Water Int.*, 25(1), 76–88, 2000.

Snyder, F. F., Synthetic Unit Hydrograph, *Trans. Am. Geophys. Union*, 19, 447–454, 1938.

Thornthwaite, C. W. and Mather, J. R., The water balance, in *Publications in Climatology: Laboratory Climatology,* Vol. 8, No. 1, Drexel Institute of Technology, Centerton, NJ, 1955.

Vetana Systems, *Vensim User's Guide*, Ventana Systems, Inc., Belmont, MA, 1995.

Viessman, J. R. and Hammer, M., *Water Supply and Pollution Control*, Harper & Row, New York, 1985.

White, I. D., Mottershead, D. N., and Harrison, S. J., *Environmental Systems: An Introductory Text*, Chapman & Hall, London, 1992.

World Resources Institute, *World Resources: A Guide to the Global Environment*, a joint publication by The World Resources Institute, The United Nations Environment Programme, The United Nations Development Programme, and The World Bank, Oxford University Press, London, 1998.

7 Groundwater Management

7.1 INTRODUCTION

The term *groundwater* usually refers to subsurface water located below the water table in saturated soils and geologic formations. In groundwater studies, an understanding of surface water and subsurface, unsaturated soil moisture, and their interactions is important. Groundwater is an important feature of the environment and a part of the hydrologic cycle; therefore, an understanding of its role in this cycle is necessary if integrated analyses are to be used in the study of watershed resources and regional assessment of contamination (Freeze and Cherry, 1979). The main objective of groundwater management is the optimal allocation of groundwater resources to water demands, taking into consideration the complex economic, environmental, and hydrogeologic constraints and conflicts in objectives. This chapter discusses the main characteristics of groundwater systems, groundwater flow equations, and groundwater modeling, and some optimization models for groundwater management are also presented.

7.2 CHARACTERISTICS OF GROUNDWATER SYSTEMS

7.2.1 Hydraulic Properties of Soils and Groundwater

Groundwater occurs whenever the soil pores are occupied and saturated by water. An *aquifer* is a permeable geologic formation that can transmit significant quantities of water under the usual hydraulic gradient. The saturated geologic unit is referred to as an *aquiclude* when it is not capable of transmitting significant volumes of water under ordinary hydraulic conductivity. An *aquitard* is permeable enough to transmit water in significant quantities within regional groundwater modeling but not in sufficient quantities for installing production wells within them. A *confined aquifer* is an aquifer that has been confined between two aquitards; however, when the water table forms the upper boundary, the aquifer is *unconfined*.

Hydraulic head at a given point (A in Figure 7.1) in groundwater is defined as follows:

$$h = Z + \frac{P}{\rho g} = Z + \psi \tag{7.1}$$

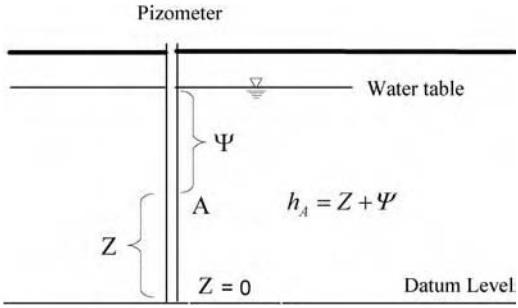


FIGURE 7.1 Illustration of hydraulic head.

As shown in Figure 7.1, h is the hydraulic head, Z is the distance between the point (A) and a datum level, P is the fluid pressure from the water column above the point, ρg (or γ) is the specific weight of water, and ψ is the pressure head.

The loss of head per unit length of flow along a streamline is the *hydraulic gradient* and is presented as:

$$i = \frac{dh}{dl} \tag{7.2}$$

The *hydraulic conductivity* is defined as the volume of water flowing perpendicular to a unit area of porous media per unit time and under unit hydraulic conductivity. The hydraulic conductivity (K) combines fluid and porous medium properties as follows:

$$K = \frac{k\rho g}{\mu} \tag{7.3}$$

where k is the *intrinsic permeability* and is a function of pore size, and μ is the *dynamic viscosity* of water, which is a function of temperature (generally decreasing when the temperature increases). The hydraulic conductivities of several lithologies are presented in Table 7.1.

Hydraulic conductivity values usually show spatial variation within a geologic porous media or with the direction of measurement, termed *heterogeneity* and *isotropy*, respectively. In other words, if the hydraulic conductivity is independent of the position in porous media, the geologic formation is *homogenous*; if it is independent of the direction of measurement in a point, the formation is *isotropic*. Water in unsaturated areas is subject to hydraulic gradient, surface tension, and molecular attraction. The unsaturated hydraulic conductivity is equal to $K\psi$, where ψ is the pressure head, which is usually negative in unsaturated media.

In the mid-19th century, Henri Darcy discovered one of the most important relationships of porous media hydrodynamics. Darcy's law defines the *specific*

TABLE 7.1
Hydraulic Conductivities of Several Lithologies

Lithology	Porosity (%)	K (cm/sec)
Unconsolidated		
Gravel	25–40	$10^{-2} - 10^{2}$
Sand	25–50	$10^{-4} - 1$
Silt	35–50	$10^{-7} - 10^{-3}$
Clay	40–70	$10^{-10} - 10^{-7}$
Glacial till	10–20	$10^{-10} - 10^{-4}$
Indurated		
Fractured basalt	5–50	$10^{-5} - 1$
Karst limestone	5–50	$10^{-4} - 10$
Sandstone	5–30	$10^{-8} - 10^{-4}$
Limestone, dolomite	0–20	$10^{-7} - 10^{-4}$

discharge (volumetric flow rate per unit area of porous media, perpendicular to the flow direction) as follows:

$$q = -K\left(\frac{dh}{dl}\right) \tag{7.4}$$

where q is the specific discharge, which has the dimension of velocity; K is the hydraulic conductivity, and dh/dl is the hydraulic gradient. The negative sign indicates the flow direction, which is equal to the direction of decreasing gradient. The average pore velocity or seepage velocity that can be used to estimate true travel time or solute transport is:

$$v = \frac{q}{n_e} \tag{7.5}$$

where n_e is the *effective porosity*, which is the fraction of pores in porous medium that allows the passage of water.

Example 7.1

As shown in Figure 7.2, a semi-impervious aquitard separates an overlying water table aquifer from an underlying confined aquifer. Determine the rate of flow that takes place between the aquifers.

Solution: Because the water table is above the piezometric surface and a semi-impervious (leaky) layer exists, flow will take place from the water table aquifer to the confined aquifer. Assume that the head at point *b* in Figure 7.2 is h_b, and consider

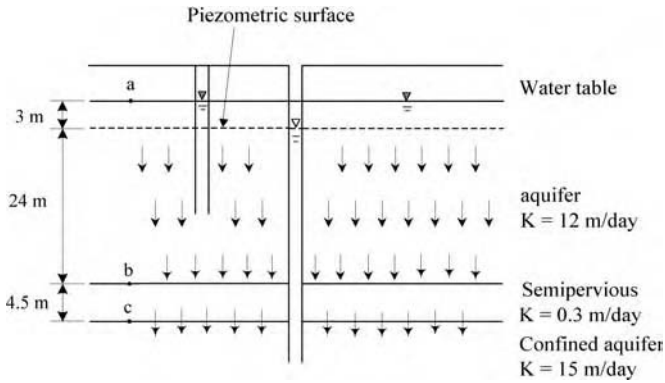


FIGURE 7.2 Vertical downward flow through soil layers for Example 7.1.

the unit horizontal area through which flow takes place. Between points *a* and *b*, from Eq. (7.4):

$$q = \frac{12(27\text{-}h_b)}{27}$$

Between points *b* and *c*, from Eq. (7.4):

$$q = (0.3)\frac{(h_b + 4.5)\text{-}(24 + 4.5)}{4.5}$$

From these equations we have:

$$h_b = 26.6\text{m}$$

$$q = 0.17\frac{\text{m}^3}{day} \text{ per square meter}$$

The *transmissivity* or *transmissibility* (*T*) in horizontal or layered aquifers is defined as:

$$T = Kb \tag{7.6}$$

where *b* is the saturated thickness of the aquifer. Therefore, *transmissivity* reflects the flow rate of water transmitted through a unit width of aquifer under a unit hydraulic gradient.

The *storativity* (or *storage coefficient*) of an aquifer is defined as the volume of water released or taken into a unit surface area per unit time and per unit decrease or increase in hydraulic head. The storativity can be defined as:

$$S = S_s b \tag{7.7}$$

where S_s is *specific storage*, and b is the saturated thickness of the aquifer. In confined aquifers, the specific storage is defined as follows:

$$S_s = \rho g(\alpha + n\beta) \tag{7.8}$$

where α and β are aquifer and water compressibility, respectively and n is soil porosity.

7.3 GROUNDWATER FLOW EQUATIONS

The mathematical description of groundwater flow is based on the conservation of mass, energy, and momentum. In this section, groundwater flow equations are developed using these principles. The groundwater equations are expressed in terms of partial differential equations, where the spatial coordinates and time are independent variables.

7.3.1 The Continuity Equation

The continuity equation as a fundamental law of groundwater flow expresses the principle of mass conservation as follows:

$$\begin{aligned}&\text{mass inflow rate} - \text{mass outflow rate}\\ &\quad = \text{rate of change of mass storage with time}\end{aligned} \tag{7.9}$$

Considering an elementary control volume of soil (Figure 7.3) that has the volume of (Δx, Δy, Δz), the mass of the groundwater (M) in this control volume is:

$$M = \rho.\theta.\Delta x.\Delta y.\Delta z \tag{7.10}$$

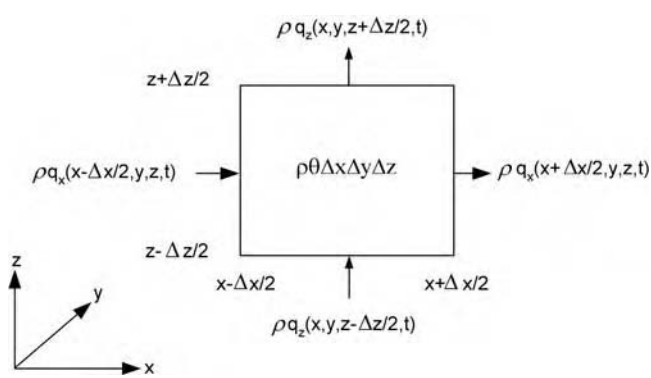


FIGURE 7.3 Mass conservation in elementary control volume (actual volume is three dimensional but it has been shown in only two dimensions for more clarity). (From Delleur, J. W., Ed., *The Handbook of Groundwater Engineering*, CRC Press/Springer-Verlag, Boca Raton, FL, 1999. With permission.)

where θ is the moisture content of the porous medium. Equation (7.9) can be rewritten as follows:

$$\frac{\partial M}{\partial t} = \text{inflow} - \text{outflow} \tag{7.11}$$

The inflow and outflow can be calculated for each side of the element. For example, considering the flux vector $\boldsymbol{q} = (q_x, q_y, q_z)$, the mass of groundwater inflow to the left side is:

$$\rho q_x(x - \Delta x/2, y, z, t)\Delta y \Delta z \approx \left(\rho q_x - \frac{\Delta x}{2}\frac{\partial \rho q_x}{\partial x}\right)\Delta y \Delta z \tag{7.12}$$

A similar equation can be derived for the other side. Considering these equations, the total inflow minus outflow can be derived as follows:

$$\frac{\partial M}{\partial t} = -\left(\frac{\partial \rho q_x}{\partial x} + \frac{\partial \rho q_y}{\partial y} + \frac{\partial \rho q_z}{\partial z}\right)\Delta x \Delta y \Delta z \tag{7.13}$$

Based on Eq. (7.11), the change in storage is given by:

$$\frac{\partial M}{\partial t} = \frac{\partial}{\partial t}(\rho\,\theta\,\Delta x\,\Delta y\,\Delta z) = \rho\left(\frac{\theta}{\rho}\frac{\partial \rho}{\partial t} + \frac{\partial \theta}{\partial t} + \frac{\theta}{\Delta z}\frac{\partial \Delta z}{\partial t}\right)\Delta x\,\Delta y\,\Delta z \tag{7.14}$$

Compression of a porous medium and water can be considered using the following equations, respectively:

$$\frac{1}{\Delta z}\frac{\partial \Delta z}{\partial t} = \alpha \frac{\partial p}{\partial t} \tag{7.15}$$

$$\frac{1}{\rho}\frac{\partial \rho}{\partial t} = \beta \frac{\partial \rho}{\partial t} \tag{7.16}$$

where α is the elastic compressibility coefficient of porous medium, p is the water pressure, and β is the compressibility coefficient of water. Substitution of these equations into Eq. (7.14) gives:

$$\frac{\partial M}{\partial t} = \rho\left(\theta(\alpha + \beta)\frac{\partial p}{\partial t} + \frac{\partial \theta}{\partial t}\right)\Delta x\,\Delta y\,\Delta z \tag{7.17}$$

Therefore, using Eqs. (7.10) and (7.13), the continuity equation can be derived as:

$$\rho\left(\theta(\alpha+\beta)\frac{\partial p}{\partial t}+\frac{\partial \theta}{\partial t}\right)=-\left(\frac{\partial \rho q_x}{\partial x}+\frac{\partial \rho q_y}{\partial y}+\frac{\partial \rho q_z}{\partial z}\right) \tag{7.18}$$

The left side of this equation describes the change in volume of water in a porous medium due to a change in water content or compression of the water and the medium.

Example 7.2

Develop the steady-state flow equation in an aquifer that has the following hydraulic conductivity function:

$$K = K_0(a + be^{cx} + de^{fy})$$

where K_0, a, b, c, d, and f are parameters of the system.

Solution: The continuity equation (Eq. (7.18)) in the steady-state condition can be simplified as follows:

$$\frac{\partial q_x}{\partial x}+\frac{\partial q_y}{\partial y}=0$$

Applying Darcy's law, we have:

$$\frac{\partial}{\partial x}(K\frac{\partial h}{\partial x})+\frac{\partial}{\partial y}(K\frac{\partial h}{\partial y})=0$$

or,

$$\frac{\partial}{\partial x}(K_0(a+be^{cx}+de^{fy})\frac{\partial h}{\partial x})+\frac{\partial}{\partial y}(K_0(a+be^{cx}+de^{fy})\frac{\partial h}{\partial y})=0$$

Finally,

$$(a+be^{cx}+de^{fy})\frac{\partial^2 h}{\partial x^2}+cbe^{cx}\frac{\partial h}{\partial x}+(a+be^{cx}+de^{fy})\frac{\partial^2 h}{\partial y^2}+fde^{cx}\frac{\partial h}{\partial y}=0$$

7.3.2 Equation of Motion in Groundwater

The equation of motion can be derived using conservation of momentum. Considering the elementary control volume (Figure 7.3), the forces that usually act on the water in the control volume are:

- Pressure forces
- Gravity forces
- Reaction forces of solids

For example, the pressure force on the left side of the control volume is as follows:

$$\theta p(x - \Delta x / 2, y, z, t)\Delta y \Delta z \tag{7.19}$$

The left side also has the same force but in the opposite direction; therefore, the resulting pressure force component in the x direction is as follows (Delleur, 1999):

$$-\frac{\partial \theta p}{\partial x}\Delta x \Delta y \Delta z \tag{7.20}$$

The other components of the pressure force on the control volume in the y and z directions can be obtained using a similar method.

The gravity force, which is equal to the total weight of water in the control volume and acts in a direction opposite of the z direction, is:

$$-\rho g \theta \Delta x \Delta y \Delta z \tag{7.21}$$

where g is the gravity constant. The reaction forces are usually defined as average body forces per unit of water volume and consist of the friction forces due to water movement and the forces that act against the water pressure. The friction and reaction forces are denoted as $\boldsymbol{r} = (r_x, r_y, r_z)$ and $\boldsymbol{f} = (f_x, f_y, f_z)$, respectively. The x component of these forces is as follows:

$$(r_x + f_x)\theta \Delta x \Delta y \Delta z \tag{7.22}$$

It can have similar components in the y and z directions.

Using the del operator $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$, the effects of all forces can be combined in one vector as follows:

$$\left[-\nabla(\theta p) - \rho g \theta \nabla z + (\boldsymbol{r} + \boldsymbol{f})\theta\right]\Delta x \Delta y \Delta z \tag{7.23}$$

Considering that when the fluid is at rest, the friction f is zero and the water pressure is hydrostatic, the overall reaction force can be evaluated as follows (Delleur, 1999):

$$\boldsymbol{r} = \frac{p}{\theta}\nabla\theta \tag{7.24}$$

In the case of groundwater motion, the sum of the forces equal to the changes of fluid momentum and the friction force is not zero. As the groundwater flow is generally very slow, the changes in momentum are negligible and the forces that act on the fluid in the control volume are approximately in equilibrium:

$$-\nabla p - \rho g \nabla z + \boldsymbol{f} \approx 0 \tag{7.25}$$

The friction force can be presented as:

$$\boldsymbol{f} = -\frac{\mu}{k}\boldsymbol{q} \tag{7.26}$$

where μ is the dynamic viscosity of the fluid, k is the intrinsic permeability, and q is the groundwater flux. The equation of motion can be expressed using Eqs. (7.20) and (7.21) as follows:

$$\boldsymbol{q} = -\frac{k}{\mu}(\nabla p + \rho g \nabla z) \tag{7.27}$$

When the gradients of density are negligible, the motion equation can be simplified as:

$$\boldsymbol{q} = -\frac{k\rho g}{\mu}(\nabla\psi + z) = -K\nabla h \tag{7.28}$$

where

$$\psi = \int \frac{dp}{\rho g}$$

is the pressure potential and the other variables have been defined in previous sections. The above equation clarifies the principles and assumptions that result in Darcy's law. One of the most important assumptions of Darcy's law is that the flow of water in a porous medium is slow and the large values of friction forces balance the driving forces. In the anisotropic porous media, Darcy's law in Cartesian coordinates becomes:

$$q_x = -K_x \frac{\partial h}{\partial x} \tag{7.29}$$

$$q_y = -K_y \frac{\partial h}{\partial y} \tag{7.30}$$

$$q_z = -K_z \frac{\partial h}{\partial z} \tag{7.31}$$

7.3.3 The Groundwater Flow Equation

The groundwater equation can be derived by a combination of the continuity equation and the equation of motion as expressed in Eqs. (7.18) and (7.28), as follows.

$$\theta(\alpha + \beta)\frac{\partial p}{\partial t} + \frac{\partial \theta}{\partial t} = \nabla\left[\frac{k}{\mu}(\nabla p + \rho g \nabla z)\right] \tag{7.32}$$

This equation can be written as follows:

$$S\frac{\partial p}{\partial t} = \nabla\left[\frac{k}{\mu}(\nabla p + \rho g \nabla z)\right] \tag{7.33}$$

where S is the storage coefficient and is equal to:

$$S = \theta(\alpha + \beta) + \frac{\partial \theta}{\partial p} \tag{7.34}$$

S is related to water and soil characteristics such as saturated and unsaturated conditions of soil and the water pressure. Therefore, the three-dimensional equation of groundwater movement as a function of water pressure (not as a function of groundwater potential) can be derived as follows:

$$S\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}\left(\frac{k_x}{\mu}\frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{k_y}{\mu}\frac{\partial p}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{k_z}{\mu}\left(\frac{\partial p}{\partial z} + \rho g\right)\right) \tag{7.35}$$

The general form of the equation of groundwater flow is usually simplified in practice. When the density effects can be ignored, Eq. (7.32) is simplified as follows:

$$\theta(\alpha + \beta)\frac{\partial p}{\partial t} + \frac{\partial \theta}{\partial t} = \nabla.(K\nabla h) \tag{7.36}$$

In groundwater, the density is usually considered to be constant; therefore, the temporal variation of water pressure and groundwater potential are related as follows:

$$\frac{\partial p}{\partial t} = \rho g \frac{\partial h}{\partial t} \tag{7.37}$$

Therefore, the following basic groundwater flow equation is derived:

$$\rho g \theta(\alpha + \beta) \frac{\partial h}{\partial t} + \frac{\partial \theta}{\partial t} = S_s \frac{\partial h}{\partial t} = \nabla.(K \nabla h) \tag{7.38}$$

where S_s is the specific storage coefficient. Written in Cartesian coordinates, the equation of saturated groundwater flow becomes:

$$S_s \frac{\partial h}{\partial t} = \frac{\partial}{\partial x}\left(K_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial h}{\partial z}\right) \tag{7.39}$$

In the steady-state condition, Eq. (7.39) can be simplified as follows:

$$\frac{\partial}{\partial x}\left(K_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial h}{\partial z}\right) = 0 \tag{7.40}$$

This equation shows that the difference in groundwater potential causes the movement of water in porous media and the fluxes depend on the hydraulic conductivity of the medium in different directions.

Example 7.3

Develop the groundwater flow equations for a two-dimensional, horizontal, semi-confined or leaky aquifer that has the thickness $b(x, y)$ at each point (x, y).

Solution: The average value of hydraulic head at each point is:

$$\bar{h} = \frac{1}{b} \int_0^b h \, dz \tag{7.41}$$

Using Eq. (7.39) and considering $T = K \cdot b$, the flow equation is:

$$S_s b \frac{\partial \bar{h}}{\partial t} = \frac{\partial}{\partial x}\left(T_x \frac{\partial \bar{h}}{\partial x}\right) + \frac{\partial}{\partial y}\left(T_y \frac{\partial \bar{h}}{\partial y}\right) + q_z(b) \tag{7.42}$$

where $q_z(b)$ is the vertical recharge or discharge, such as leakage, pumping, or injection. The vertical leakage from the upper semiconfining layer can be calculated using Darcy's law as follows:

$$q_z(b) = \frac{K_a(H_a - \bar{h})}{b_a} \tag{7.43}$$

where K_a is the hydraulic conductivity of the upper confining layer, H_a is the head in the upper boundary of the confining layer, and b_a is the thickness of the semiconfined layer. Considering the point injection and pumping wells in the system, the flow equation can be written as follows (Hantush and Jacob, 1955; Willis and Yeh, 1987):

$$S_s b \frac{\partial \bar{h}}{\partial t} = \frac{\partial}{\partial x}\left(T_x \frac{\partial \bar{h}}{\partial x}\right) + \frac{\partial}{\partial y}\left(T_y \frac{\partial \bar{h}}{\partial y}\right) + \frac{K_a(H_a - \bar{h})}{b_a} \pm \sum_{w \in W} Q_w \delta_w \tag{7.44}$$

where $-Q_w$ is the discharge ($+Q_w$ is the recharge) from pumping (injection) well w, and δ is a 0/1 integer variable. The variable $\delta = 1$ for pumping or injection at well site w.

7.4 GROUNDWATER MODELING

Prediction of subsurface flow, water table level, solute transport, and simulation of natural or human-induced stresses is necessary for groundwater management. Groundwater simulation models have been widely used in groundwater systems analysis and management. These models generally require the solution of partial differential equation. Analytical models can yield the exact solutions for some simple or idealized problems. Numerical models provide approximate solutions by discretization of time and space, and because they do not require the rigid idealized conditions of analytical models they are more flexible. In this section, some basic aspects of numerical groundwater modeling such as principles of numerical methods, grid generation, and model calibration and validation, are discussed.

7.4.1 The Flow Net

In a two-dimensional cross-section of a porous medium, the set of equipotential lines and flow lines is referred to as the *flow net.* Figure 7.4 shows the orthogonal network of flow at a point. In this figure, ΔS denotes the distance between two adjacent equipotential lines and ΔW is the distance between a pair of adjacent streamlines. Flow nets can be used in graphical modeling of flow in porous media.

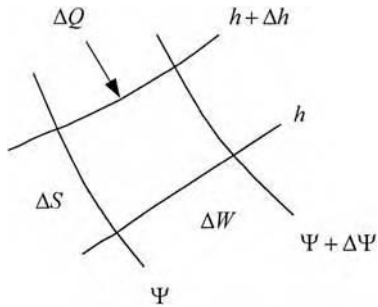


FIGURE 7.4 Flow at a point.

Delleur (1999) recommended the following procedure for drawing the flow net:

1. The boundaries of the flow region should be drawn to a scale such that the streamlines and equipotential lines terminate on the figure.
2. Three or four streamlines from infinite number of possible curves should be sketched. They must provide a smooth transition between the boundary streamlines.
3. The equivalent lines that intersect all streamlines at a right angle are drawn; therefore, the enclosed figures must form curve-linear rectangles that have the same ratio of $\Delta W/\Delta S$ along the flow channel.

Most of the analytical and graphical methods for solution of two-dimensional groundwater problems are concerned with the determination of a suitable function that can transform a problem from a geometrical domain into a domain with a more straightforward solution algorithm. A transformation that possesses the property of preserving angles of intersections and the approximate image of small shapes is *conformal mapping*. Harr (1962) presented various functions and discussed the manner in which these functions transform geometric figures from one complex plan to another. He also considered the more general problem of finding a functional relationship that will provide a specific transformation.

Example 7.4

As shown in Figure 7.5, a dam is constructed on a permeable stratum underlain by an impermeable rock. A row of sheet pile is installed at the upstream face. The hydraulic conductivity of the soil is 4.5 m/day. Determine (a) the rate of flow from upstream to downstream, and (b) the uplift pressure acting in point H.

Solution: The flow net is drawn as shown in Figure 7.5. In this figure, the number of flow channels (n_f) is equal to 5 and the number of equipotential drops (n_d) is equal to 17.

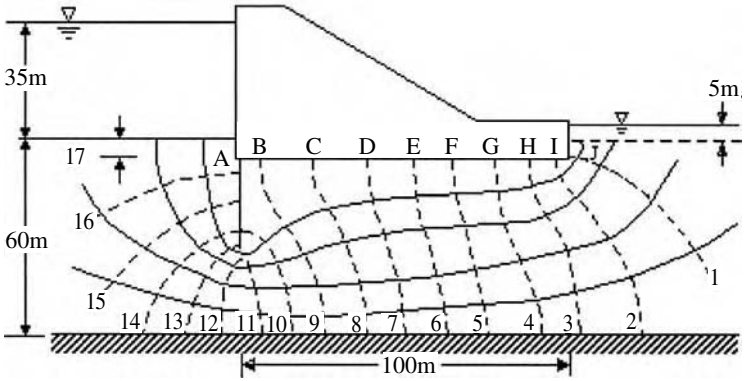


FIGURE 7.5 Net flow for a dam with sheet pile.

The flow rate or seepage per unit can be calculated using the following equation:

$$q = \frac{n_f}{n_d} Kh \tag{7.45}$$

where h is the total head loss in the flow system:

$$q = \frac{5}{17}(4.5)(30) = 39.7 \frac{\text{m}^3}{\text{day}}$$

The uplift pressure at each point can be calculated using following equation:

$$u = \left(\frac{n}{n_d} h + z \right) \gamma_w \tag{7.46}$$

where n is the number of equipotential lines, counting the last line on the downstream as 0; z is the depth of base below the datum (if the base is above the datum, z is negative), and γ_w is the specific weight of water.

$$u = \left(\frac{3}{17} \times 30 + 2 \right) \times 1 = 7.29 \frac{kN}{m^2}$$

7.4.2 Numerical Methods

Numerical modeling of groundwater was not extensively pursued until the mid-1960s. Since that time, numerical models have been developed and widely used in groundwater resources management. Numerical methods transform partial differential equations of groundwater into a set of ordinary differential or algebraic equations. State variables at discrete nodal points are determined by solving these equations.

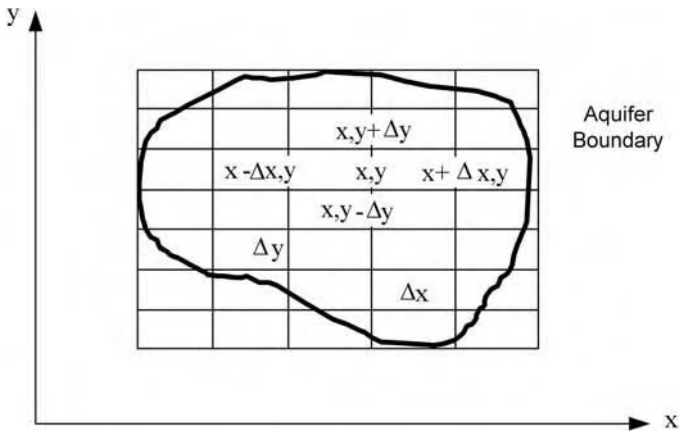


FIGURE 7.6 Finite-difference grid.

Two major classes of numerical methods are acceptable for solving groundwater problems: finite difference and finite element. In both methods, the study area is discretized by grid generation into a number of subareas called *elements* or *cells*. Figure 7.6 shows a simple discretized aquifer.

In the finite-difference method, the first derivatives in partial equations are approximated by the difference between values of independent variables at adjacent nodes, considering the distance between the nodes, and at two successive time levels, considering the duration of the time-step increment. In finite-element methods, functions of dependent variables and parameters are used to evaluate equivalent integral formulation of partial differential equations (Delleur, 1999). Although each approach has some advantages and disadvantages, the finite-difference methods are generally easier to program because of their conceptual and mathematical simplicity.

A major advantage of the finite-element method is the flexibility of this method in close spatial approximation of irregular boundaries of aquifer and parameters of zones within the aquifer. However, mesh generation and specification, and construction of input data sets for an irregular finite-element grid is much more difficult in comparison with a regular rectangular finite-difference grid.

Figure 7.7a shows a hypothetical aquifer with impermeable boundaries and a water supply well. This aquifer can be discretized using finite-difference (Figure 7.7b) and finite-element (Figure 7.7c) grids. Comparing Figures 7.7b and 7.7c shows that finer mesh can be allocated to the areas of interest to obtain a more precise solution. As shown in Figure 7.7b, in the finite-difference method with a rectangular grid, the boundaries are approximated in a stepwise manner, but finite-difference grids using triangular elements can closely follow the aquifer boundary.

7.4.2.1 Finite-Difference Numerical Model Basics

Finite-difference models of groundwater flow approximate the derivatives of continuous variables in time and space by discrete linear changes over discrete intervals

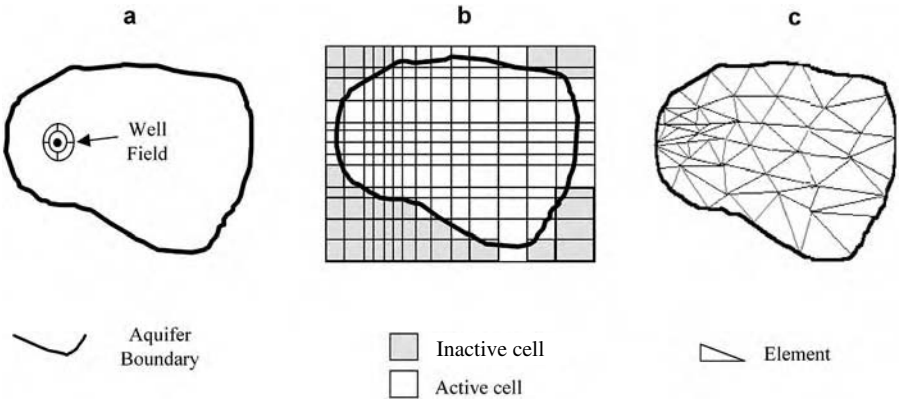


FIGURE 7.7 Application of finite-difference and finite-element grids for an irregularly bounded aquifer with a well field.

of time or space. For example, forward, backward, and central first-derivative approximations of head h are:

$$\frac{\partial h}{\partial x}\Big|_x \approx \frac{h(x+\Delta x)-h(x)}{\Delta x} + O\left((\Delta x)^2\right) \tag{7.47}$$

$$\frac{\partial h}{\partial x}\Big|_x \approx \frac{h(x)-h(x-\Delta x)}{\Delta x} + O\left((\Delta x)^2\right) \tag{7.48}$$

$$\frac{\partial h}{\partial x}\Big|_x \approx \frac{h(x+\Delta x)-h(x-\Delta x)}{2\Delta x} + O\left((\Delta x)^3\right) \tag{7.49}$$

where $O\left((\Delta x)^n\right)$ is the truncation error with an order on n.

Similarly, a reasonable approximation for the second derivative $\frac{\partial^2 h}{\partial x^2}$ can be given as:

$$\frac{\partial^2 h}{\partial x^2}\Big|_x \approx \frac{h(x+\Delta x)-2h(x)+h(x-\Delta x)}{\Delta x^2} + O\left((\Delta x)^4\right) \tag{7.50}$$

The errors of these approximations (*truncation errors*) will generally be decreased when the values of Δx and Δt are decreased.

Similar discretization should be considered for time intervals; therefore, the approximations of $\partial h/\partial t$ using the forward and backward differences are:

$$\frac{\partial h}{\partial t}\Big|_t = \frac{h^{t+1}-h^t}{\Delta t} \tag{7.51}$$

$$\frac{\partial h}{\partial t}\Big|_t = \frac{h^t-h^{t-1}}{\Delta t} \tag{7.52}$$

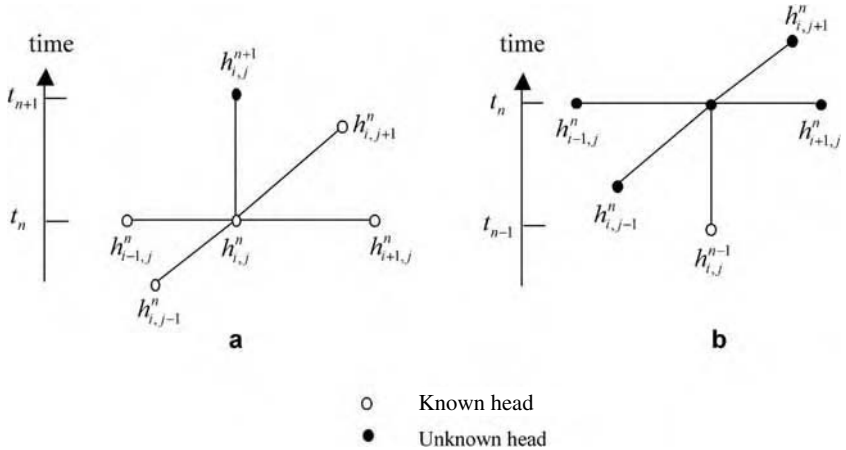


FIGURE 7.8 (a) Explicit (forward difference) and (b) implicit (backward difference) discretization of the time at node (i,j) in a two-dimensional finite-difference grid.

Figure 7.8 shows time and space discretization at node (i, j) in a two-dimensional, finite-difference grid and illustrates the application of two simple finite-difference methods for solving the groundwater flow equation. As shown in Figure 7.8a, all values of head h are known at all spatial nodes at time level n. The head at point (i, j) at time step $(n + 1)$ can be calculated using the value of h at time step n and the forward-difference time derivation. Therefore, a finite-difference equation exists for each node at time step $(n + 1)$ with only one unknown variable.

For example, in a two-dimensional groundwater flow equation for a heterogeneous, anisotropic aquifer, Eq. (7.39) can be written as:

$$S_s\left(\frac{h^{n+1}_{i,j} - h^n_{i,j}}{\Delta t}\right) = K_{x(i-\frac{1}{2},j)}\left(\frac{h^n_{i-1,j} - h^n_{i,j}}{(\Delta x)^2}\right) + K_{x(i+\frac{1}{2},j)}\left(\frac{h^n_{i+1,j} - h^n_{i,j}}{(\Delta x)^2}\right) +$$

$$K_{y(i,j-\frac{1}{2})}\left(\frac{h^n_{i,j-1} - h^n_{i,j}}{(\Delta y)^2}\right) + K_{y(i,j+\frac{1}{2})}\left(\frac{h^n_{i,j+1} - h^n_{i,j}}{(\Delta y)^2}\right) \tag{7.53}$$

In this equation, only $h^{n+1}_{i,j}$ is unknown and can be solved explicitly. Explicit finite-difference equations are simple to solve but, when time increments are too large, small numerical errors can result in larger errors in the next computational stages.

Figure 7.8b shows the time derivative as a backward difference from the head at time level $(n - 1)$, which are the known heads. Therefore, the difference equation of each node will have five unknown variables, and for a grid that has N nodes a system of N equations contains N unknown variables. This system of equations can be solved simultaneously considering the boundary conditions.

The implicit finite-difference form of the two-dimensional groundwater equation (Eq. (7.39)) is:

$$S_s\left(\frac{h_{i,j}^{n}-h_{i,j}^{n-1}}{\Delta t}\right)=K_{x(i-1/2,j)}\left(\frac{h_{i-1,j}^{n}-h_{i,j}^{n}}{(\Delta x)^2}\right)+K_{x(i+1/2,j)}\left(\frac{h_{i+1,j}^{n}-h_{i,j}^{n}}{(\Delta x)^2}\right)+$$
$$K_{y(i,j-1/2)}\left(\frac{h_{i,j-1}^{n}-h_{i,j}^{n}}{(\Delta y)^2}\right)K_{y(i,j+1/2)}\left(\frac{h_{i,j+1}^{n}-h_{i,j}^{n}}{(\Delta y)^2}\right) \tag{7.54}$$

Example 7.5

Develop the finite-difference approximation of the flow equation for a two-dimensional, semiconfined, nonhomogeneous, anisotropic aquifer.

Solution: The flow equation in the semiconfined aquifer is given by Eq. (7.44). The finite-difference approximation of this equation in a nonhomogeneous and isotropic aquifer can be as follows:

$$\frac{1}{\Delta x^2}\{T_x^{i+1/2,j}\ (h_{i+1,j}-h_{i,j})-T_x^{i-1/2,j}(h_{i,j}-h_{i-1,j})\}+\frac{1}{\Delta y^2}\{T_y^{i,j+1/2}(h_{i,j+1}-h_{i,j})$$
$$-T_y^{i,j-1/2}(h_{i,j}-h_{i,j-1})\}\pm\frac{Q_{w,i,j}}{\Delta x\,\Delta y}+\frac{K_{a,i,j}(H_{a,i,j}-h_{i,j})}{b_a}=S_{i,j}h_{i,j} \tag{7.55}$$

By developing Eq. (7.55) for all internal nodes (i, j; $i = 1, \ldots, n$ and $j = 1, \ldots, m$) of the domain, a set of linear equations can be expressed in the following form:

$$A\boldsymbol{h}^{\bullet}+B\boldsymbol{h}+\boldsymbol{g}=0 \tag{7.56}$$

In these equations, which are referred to as *dynamic response equations*, $\boldsymbol{h}$ is a column vector of unknown heads, where $h = (h_{11}, \ldots, h_{1m}, \ldots, h_{n,1}, \ldots, h_{nm})^T$ and $\boldsymbol{h}^{\bullet}$ is the temporal derivate of vector $\boldsymbol{h}$. The coefficient matrices A and B depend on the flow and hydraulic properties of the porous medium. In Eq. (7.54), vector $\boldsymbol{g}$ contains the rate of pumping and injections that can be the decision variables in the groundwater planning models. Dynamic response equations can easily provide the hydraulic head at the nodes as a linear function of initial condition and the pumping/injection rate of well sites.

7.4.2.1.1 Alternating Direction Implicit (ADI) Method

Considering that the first space derivative of the head is calculated at time level $(n + 1)$ and the second at the current time level (n), the groundwater flow equation can be written as follows:

$$\frac{h_{i-1,j}^{n+1}-2h_{i,j}^{n+1}+h_{i+1,j}^{n+1}}{\Delta x^2}+\frac{h_{i,j-1}^{n}-2h_{i,j}^{n}+h_{i,j+1}^{n}}{\Delta y^2}=\frac{S}{T}\frac{h_{i,j}^{n+1}-h_{i,j}^{n}}{\Delta t}+\frac{q_{i,j}^{n}}{T} \tag{7.57}$$

Three unknown values exist at time level $(n + 1)$ and other values at time level (n) are known. Equation (7.57) can be rewritten as:

$$h_{i-1,j}^{n+1} + B_i h_{i,j}^{n+1} + h_{i+1,j}^{n+1} = D_i \tag{7.58}$$

where B_i and D_i are constant and can be calculated as follows:

$$B_i = -\left(2 + \frac{S}{T}\frac{\Delta x^2}{\Delta t}\right) \tag{7.59}$$

$$D_i = -h_{i,j-1}^{n}\frac{\Delta x^2}{\Delta y^2} - h_{i,j}^{n}\left(\frac{S}{T}\frac{\Delta x^2}{\Delta t} - \frac{2\Delta x^2}{\Delta y^2}\right) - h_{i,j+1}^{n}\frac{\Delta x^2}{\Delta y^2} + \frac{q_{i,j}^{n}}{T}\Delta x^2 \tag{7.60}$$

If Eq. (7.58) is applied for all nodes along the x-axis (jth column), we will have $(k - 2)$ simultaneous equations because the value of the head at the two outer nodes is known from the boundary conditions (k is the number of nodes along the x-axis). Therefore, by solving these equations it is possible to calculate the heads on a line parallel to the x-axis. The coefficient matrix of this set of equations is tridiagonal.

In the next time step, $\partial^2 h/\partial y^2$ is approximated at time level $(n + 2)$, while $\partial^2 h/\partial x^2$ is retained at the current level of $(n + 1)$; thus (Gupta, 1989),

$$\frac{h_{i-1,j}^{n+1} - 2h_{i,j}^{n+1} + h_{i+1,j}^{n+1}}{\Delta x^2} + \frac{h_{i,j-1}^{n+2} - 2h_{i,j}^{n+2} + h_{i,j+1}^{n+2}}{\Delta y^2} = \frac{S}{T}\frac{h_{i,j}^{n+2} - h_{i,j}^{n+1}}{\Delta t} + \frac{q_{i,j}^{n+1}}{T} \tag{7.61}$$

Three unknown values exist at time level $(n + 2)$ and other values at time level $(n + 1)$ are known. Equation (7.61) can be rewritten as follows:

$$h_{i,j-1}^{n+2} + B_j h_{i,j}^{n+2} + h_{i,j+1}^{n+2} = D_j \tag{7.62}$$

where:

$$B_i = -\left(2 + \frac{S}{T}\frac{\Delta y^2}{\Delta t}\right) \tag{7.63}$$

$$D_i = -h_{i-1,j}^{n+1}\frac{\Delta y^2}{\Delta x^2} - h_{i,j}^{n+1}\left(\frac{S}{T}\frac{\Delta y^2}{\Delta t} - \frac{2\Delta y^2}{\Delta x^2}\right) - h_{i+1,j}^{n+1}\frac{\Delta y^2}{\Delta x^2} + \frac{q_{i,j}^{n+1}}{T}\Delta y^2 \tag{7.64}$$

Applying Eq. (7.62) for all nodes along the y-axis (ith column), we will have $(m - 2)$ simultaneous equations because the value of the head at the two outer nodes is known from the boundary conditions (m is the number of nodes along the y-axis). Therefore, by solving $(m - 2)$ equations it is possible to calculate the heads of a line parallel to the y-axis.

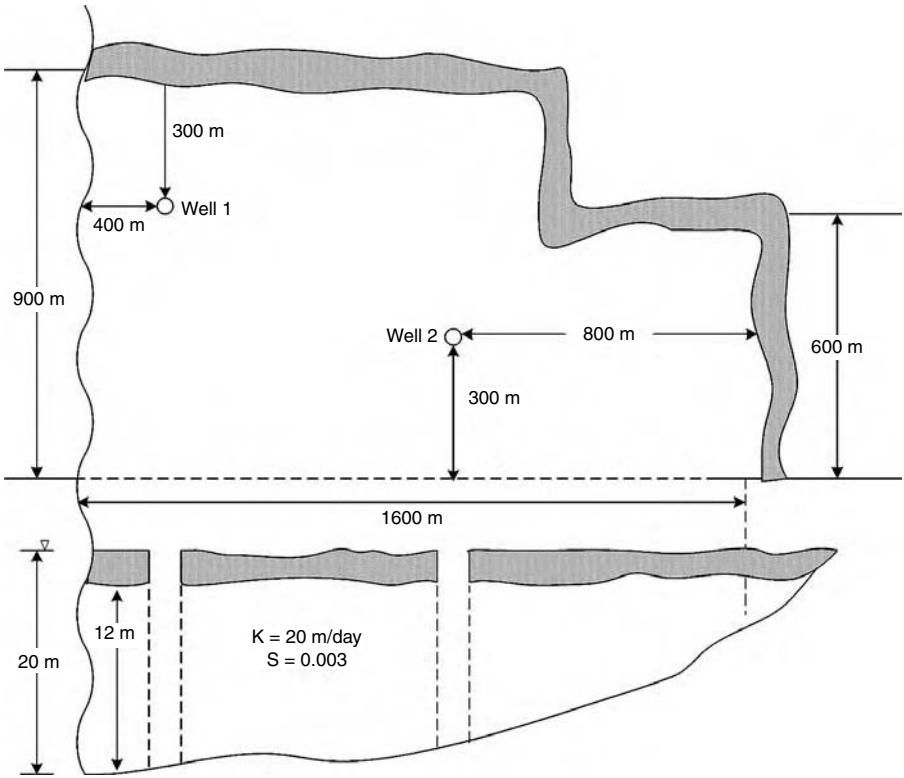


FIGURE 7.9 Plan of confined aquifer and its cross-section. (Example 7.6)

Example 7.6

Two fully penetrated wells in a confined aquifer system are shown in Figure 7.9. Wells 1 and 2 pump at a rate of 90 L/sec and 60 L/sec, respectively. The initial piezometric head is horizontal, in level with the stream. Determine the spatial variation of the piezometric head after one day of pumping. The aquifer is bounded by a stream on one side and impermeable boundaries on two sides. It is extended semi-infinitely on the remaining side, but a width of 900 m is used for the flow system under consideration.

Solution: The grid layout is shown in Figure 7.10. In the initial conditions, the heads of all nodes are 20 m at $t = 0$, and boundary conditions are:

Recharge boundary:

$$h_{1,1} = h_{1,2} = h_{1,3} = h_{1,4} = 20 \text{ m}$$

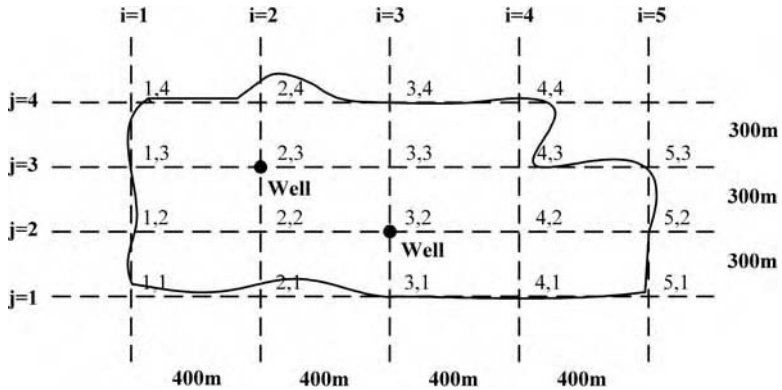


FIGURE 7.10 Finite-difference grid and its coordinates of Example 7.6.

Impermeable boundaries

$$h_{5,1} = h_{4,1} \qquad h_{2,4} = h_{2,3}$$
$$h_{5,2} = h_{4,2} \qquad h_{3,4} = h_{3,3}$$
$$h_{4,3} = h_{3,3} \qquad h_{4,4} = h_{4,3}$$
$$h_{4,4} = h_{3,4}$$

$$T = \mathrm{K} \times b = 20 \times 12 = 240 \text{ m}^2/\text{day}$$

First time step: A time step Δt of 0.5 day is selected as the time interval. (Maximum time step is selected based on stability requirement; see Gupta [1989] for more details.) Because there is no pumping on the first row, piezometric heads in this row will not change in the first time step. For row 2 ($j = 2$), we have:

$$B_2 = -(2 + \frac{S}{T} \times \frac{(\Delta x)^2}{\Delta t}) = -(2 + \frac{0.003}{240} \times \frac{(400)^2}{0.5}) = -6.0$$
$$B_2 = B_3 = B_4 = -6.0$$

and

$$D_2 = -20 \times \frac{(400)^2}{(300)^2} - 20\left(\frac{0.003}{240} \times \frac{\left(400^2\right)}{0.5} - 2 \times \frac{400^2}{300^2}\right) - 20 \times \frac{400^2}{300^2} = -80$$

Since

$$q = \frac{0.06}{300 \times 400} = 0.5 \times 10^{-6} \quad \text{m/s} = 0.0432 \text{ m/day}$$

then

$$D_3 = -80 + \frac{0.0432\,(400)^2}{240} = -51.2$$

$$D_4 = D_2 = -80$$

For node 2,

$$h_{1,2}^{0.5} + B_2 h_{2,2}^{0.5} + h_{3,2}^{0.5} = D_2$$

For node 3,

$$h_{2,2}^{0.5} + B_3 h_{3,2}^{0.5} + h_{4,2}^{0.5} = D_3$$

For node 4,

$$h_{3,2}^{0.5} + B_4 h_{4,2}^{0.5} + h_{5,2}^{0.5} = D_4$$

Substituting the B, D, and initial and boundary conditions, we have:

$$20 - 6\,h_{2,2}^{0.5} + h_{3,2}^{0.5} = -80$$

$$h_{2,2}^{0.5} - 6\,h_{3,2}^{0.5} + h_{4,2}^{0.5} = -51.2$$

$$h_{3,2}^{0.5} - 6\,h_{4,2}^{0.5} + h_{5,2}^{0.5} = -80$$

or

$$\begin{bmatrix} -6 & 1 & 0 \\ 1 & -6 & 1 \\ 0 & 1 & -5 \end{bmatrix} \begin{bmatrix} h_{2,2}^{0.5} \\ h_{3,2}^{0.5} \\ h_{4,2}^{0.5} \end{bmatrix} = \begin{bmatrix} -100 \\ -51.2 \\ -80 \end{bmatrix}$$

Solving gives

$$h_{2,2}^{0.5} = 19.15 \text{ m}$$

$$h_{3,2}^{0.5} = 14.89 \text{ m}$$

$$h_{4,2}^{0.5} = 18.97 \text{ m}$$

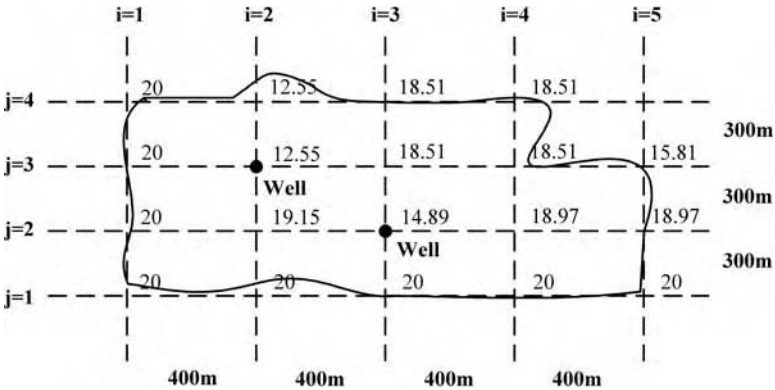


FIGURE 7.11 Heads at the end of the first time step (at $t = 0.5$ day).

The value of the heads of nodes for row 3 is solved by a similar procedure. Row 4 is an impermeable boundary; therefore, it will have the same heads as in row 3. The value of nodal heads after the first time step is shown in Figure 7.11.

Second time step: The results of the first time step are used as the initial heads of the second time step. In this time step, computations are made column by column. For column 2 ($i = 2$):

$$B_1 = -(2 + \frac{0.003}{240} \times \frac{(400)^2}{0.5}) = -4.25$$

$$B_1 = B_2 = B_3 = -4.25$$

and

$$D_1 = -20 \times \frac{(300)^2}{(400)^2} - 20\left(\frac{0.003}{240} \times \frac{\left(300^2\right)}{0.5} - 2 \times \frac{300^2}{400^2} \right) - 20 \times \frac{300^2}{400^2} = -45$$

$$D_2 = -20 \times \frac{(300)^2}{(400)^2} - 19.15\left(\frac{0.003}{240} \times \frac{\left(300^2\right)}{0.5} - 2 \times \frac{300^2}{400^2} \right) - 14.89 \times \frac{300^2}{400^2} = -41.17$$

Because

$$q = \frac{0.09}{300 \times 400} = 0.75 \times 10^{-6} \quad \text{m/s} = 0.0648 \text{ m/day}$$

then

$$D_3 = -20 \times \frac{(300)^2}{(400)^2} - 12.55\left(\frac{0.003}{240} \times \frac{\left(300^2\right)}{0.5} - 2 \times \frac{300^2}{400^2}\right)$$

$$-18.51 \times \frac{300^2}{400^2} + \frac{0.0648\,(300)^2}{240} = -11.48$$

For node 1,

$$h^1_{2,0} + B_1 h^1_{2,1} + h^1_{2,2} = D_1$$

As the aquifer in this direction extends beyond the grid boundary, the head outside the boundary is included and is assumed to be equal to the initial head of 20 m.

For node 2,

$$h^1_{2,1} + B_2 h^2_{2,2} + h^1_{2,3} = D_2$$

For node 3,

$$h^1_{2,2} + B_3 h^1_{2,3} + h^1_{2,4} = D_3$$

Substituting the B, D, and initial and boundary conditions, we have:

$$20 - 4.25\, h^1_{2,1} + h^1_{2,2} = -45$$

$$h^1_{2,1} - 4.25\, h^1_{2,2} + h^1_{2,3} = -41.17$$

$$h^1_{2,2} - 4.25\, h^1_{2,3} + h^1_{2,3} = -11.48$$

or

$$\begin{bmatrix} -4.25 & 1 & 0 \\ 1 & -4.25 & 1 \\ 0 & 1 & -3.25 \end{bmatrix} \begin{bmatrix} h^1_{2,1} \\ h^1_{2,2} \\ h^1_{2,3} \end{bmatrix} = \begin{bmatrix} -65 \\ -41.17 \\ -11.48 \end{bmatrix}$$

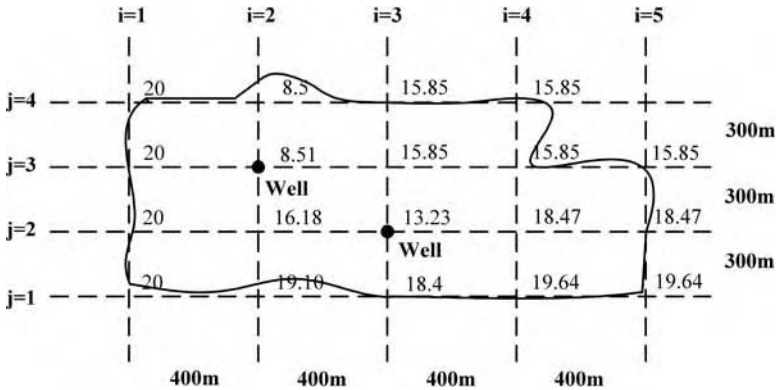


FIGURE 7.12 Heads at the end of the second time step (at t = 1 day).

Solving gives

$$h_{2,1}^{1} = 19.10 \text{ m}$$

$$h_{2,2}^{1} = 16.18 \text{ m}$$

$$h_{2,3}^{1} = 8.51 \text{ m}$$

The heads of the nodes in columns are solved by a similar procedure. The values of the nodal heads after the second time step (after t = 1 day) are shown in Figure 7.12.

7.4.2.2 Finite-Element Methods

The *finite-element* method is another numerical analysis technique for approximate solutions to groundwater flow problems. The heterogeneity of porous media and irregular boundary conditions can be handled by these methods, in which the domain is discretized into a set of elements of different sizes or shapes. Several approaches such as direct, weighted residual, and variational, have been proposed to formulate the finite-element method for a problem.

In the weighted residual method, which is frequently used for groundwater problems, the trial solution for the system of equations is written as a finite series. If the equations of the groundwater system are expressed in terms of differential operator L as follows:

$$L(h) = 0 \tag{7.65}$$

then the approximate solution in a one-dimensional problem can be written as follows:

$$h \approx \hat{h}(x,t) = \sum_{i=1}^{n} N_i(x)\, \tilde{h}_i(t) \tag{7.66}$$

where $\hat{h}$ is the approximate solution, and $N_i(x)$ are the basis or shape functions of node i defined over the entire domain. Shape functions are defined based on assumptions regarding the spatial variation of the state variables in the domain; n is the number of linearly independent shape functions, and $\tilde{h}_i(t)$ are the unknown coefficients (heads) that are determined for each node of the finite-element grid. The residual error, R, can be estimated by substituting the approximate solution $\hat{h}$ into Eq. (7.65) as follows:

$$R = L(\hat{h}) \neq 0 \tag{7.67}$$

The unknown coefficients $\tilde{h}_i$ are determined by minimizing the residual error of the approximate solution. This is accomplished by integrating the weighted error over the domain (D) and setting it equal to zero as follows:

$$\int_D W_k R\, dD = \int_D W_k L(\hat{h}) dD = 0 \quad k = 1, 2, \ldots, n \tag{7.68}$$

where W_k is the weighting function for node i. To determine the n state variables at nodal points, the weighting function should be specified and the integral equation is broken into n simple equations.

The type of weighted residual method is determined by the choice of the n weighting functions. In the Galerkin method, the weighting functions are selected to be identical to the shape functions. The modified integral equation is then given by:

$$\int_D N_k\ L(\hat{h})\, dD = 0 \quad k = 1, 2, \ldots, n \tag{7.69}$$

The modified integral equation is straightforward to solve and can be converted to n simultaneous algebraic equations. The following example illustrates the application of Galerkin's method to develop the dynamic response equation for a semiconfined aquifer system. For a more comprehensive explanation of this method, see Cooley (1992).

Example 7.7

Develop the dynamic response equation for a two-dimensional semiconfined, nonhomogeneous, anisotropic aquifer using the Galerkin procedure.

Solution: Equation (7.69) can be rewritten considering all elements of the system as follows:

$$\sum_e \int_{D_e} N_k \; L(\hat{h})\, dD_e = 0 \qquad \forall k \tag{7.70}$$

where D_e is the elemental domain and $L(\hat{h})$ can be represented by Eq. (7.44). Considering the equation $\hat{\boldsymbol{h}} = \boldsymbol{N}\tilde{\boldsymbol{h}}$, Eq. (7.70) can be rewritten as follows:

$$\sum_e \int_{D_e} N_k \; L(\boldsymbol{N}\tilde{\boldsymbol{h}})\, dD_e = 0 \qquad \forall k \tag{7.71}$$

where $\tilde{\boldsymbol{h}}$ is the vector of unknown nodal head values and N is global shape function which is a matrix of the nodal shape function. By substituting Eq. (7.44) in Eq. (7.71), we have:

$$\sum_e \int_{D_e} \left\{ \nabla.\mathrm{T}\nabla(\boldsymbol{N}\tilde{\boldsymbol{h}}) + \frac{K_a}{b_a} H_a - \frac{K_a}{b_a} \boldsymbol{N}\tilde{\boldsymbol{h}} - S\frac{\partial}{\partial t}(\boldsymbol{N}\tilde{\boldsymbol{h}}) \pm Q \right\} N_k dD_e = 0 \qquad \forall k \tag{7.72}$$

where Q indicates the magnitude of the source or sink. For an element, each term of this integral equation produces an elemental matrix that is dependent on the system parameters and shape (basis) functions. For example, the fourth term for each element can be written as follows:

$$-SA_e\tilde{\boldsymbol{h}}^{\bullet} \tag{7.73}$$

where $\tilde{\boldsymbol{h}}^{\bullet} = [\tilde{h}_i^{\bullet}, \tilde{h}_j^{\bullet}, ..., \tilde{h}_k^{\bullet}]^T$, and matrix A_e is:

$$A_e = \begin{bmatrix} \int N_i^2 dD_e & \int N_i N_j dD_e & \cdots & \int N_i N_k dD_e \\ \int N_j N_i dD_e & \int N_j^2 dD_e & \cdots & \\ \vdots & \vdots & & \\ \int N_k N_i dD_e & \int N_k N_j dD_e & \cdots & \int N_k^2 dD_e \end{bmatrix} \tag{7.74}$$

Combining the various matrix terms and dropping the ~ notation, Eq. (7.72) can be expressed as follows (Willis and Yeh, 1987):

$$\boldsymbol{A}\boldsymbol{h}^{\bullet} + \boldsymbol{B}\boldsymbol{h} + \boldsymbol{g} = \boldsymbol{0} \tag{7.75}$$

The initial condition can be $\boldsymbol{h}(0) = g(x)$.

For more information about finite-element methods such as global matrix generation, formulation of general basis functions, and solution methods for dynamic response equations, the reader is referred to Zienkiewicz (1971) and Willis and Yeh (1987).

Example 7.8

Solve the differential equation of $d^2h/dx^2 - h = 0$ using the Galerkin method and considering the following conditions:

$$0 \le x \le 3$$
$$h = 0 \text{ m} \qquad \text{where} \qquad x = 0$$
$$h = 10 \text{ m} \qquad \text{where} \qquad x = 3$$

Solution: As shown in Figure 7.13, a set of N nodes and M elements is selected; for each node, a partially linear-shaped function $N_i(x)$ is defined. Therefore, the approximate function $\hat{h}(x, t)$ at node i is defined as follows:

$$h \approx \hat{h}(x,t) = \sum_{i=1}^{w} N_i(x)\hat{h}_i(t) \qquad 0 \le x \le 3$$

Considering the unknown variables (h_1, h_2, …, h_N), Eq. (7.68) can be rewritten as follows:

$$\int_0^3 W_i\left(\frac{d^2\hat{h}}{dx^2} - \hat{h}\right)dx = 0 \qquad i = 1,2,\ldots,N \tag{7.76}$$

or

$$\int_0^3 \left(\frac{dW_i}{dx}\cdot\frac{d\hat{h}}{dx} + W_i\hat{h}\right)dx - \left[W_i\frac{d\hat{h}}{dx}\right] = 0 \qquad t = 1,2,\ldots,N \tag{7.77}$$

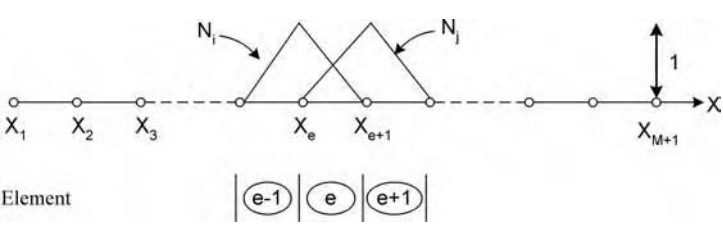


FIGURE 7.13 General numbering of nodes and elements.

Equation (7.77) can be rewritten as follows using the Galerkin method ($W_i = N_i$ and $\hat{\boldsymbol{h}} = \boldsymbol{N} \cdot \boldsymbol{h}$):

$$\boldsymbol{A}\boldsymbol{h} = \boldsymbol{g} \tag{7.78}$$

$$A_{ij} = \int_0^3 \left(\frac{dN_i}{dx} \cdot \frac{dN_j}{dx} + N_i N_j\right) dx \qquad 1 \le i, j \le N \tag{7.79}$$

$$g_i = \left[N_i \frac{dh}{dx}\right]_0^3 \qquad 1 \le i \le N \tag{7.80}$$

If the above equation is used for each element e between nodes m and n, the coefficients of the equation can be calculated using Eq. (7.71). Assuming $X = x - x_m$, the shape functions are as follows:

$$N_m = N_m^e = \frac{X}{L^e} \tag{7.81}$$

$$N_n = N_n^e = \frac{h^e - X}{L^e} \tag{7.82}$$

where $L^e = x_n - x_m$. If node k is not related to element e, N_k will be zero for element e. Therefore, the components of matrix $\boldsymbol{A}$ are calculated as follows:

$$A_{mn}^e = A_{nm}^e = \int_0^{L^e} \left[\frac{dN_m^e}{dx} \cdot \frac{dN_n^e}{dx} + N_m^e N_n^e\right] dx = \frac{-1}{L^e} + \frac{L^e}{6}$$

$$A_{mm}^e = A_{nn}^e = \int_0^{L^e} \left[\left(\frac{dN_m^e}{dx}\right)^2 + \left(N_n^e\right)^2\right] dx = \frac{1}{L^e} + \frac{L^e}{3}$$

Matrix $\boldsymbol{A}$ is calculated from the above submatrices $\boldsymbol{A}^e$, considering the position of the element in Figure 7.14.

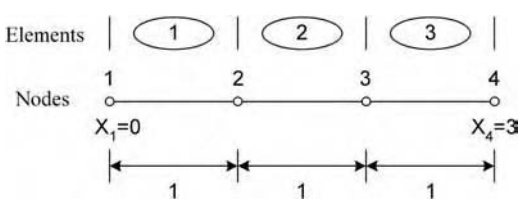


FIGURE 7.14 Numbering of nodes and elements for Example 7.8.

Assuming $L_1 = L_2 = L_3 = L$, the matrices $\boldsymbol{A}^e$ (e = 1,2,3) and $\boldsymbol{A}$ can be calculated as:

$$\boldsymbol{A}^1 = \begin{bmatrix} \frac{1}{L}+\frac{L}{3} & \frac{-1}{L}+\frac{L}{6} & 0 & 0 \\ \frac{-1}{L}+\frac{L}{6} & \frac{1}{L}+\frac{L}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{A}^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{L}+\frac{L}{3} & \frac{-1}{L}+\frac{L}{6} & 0 \\ 0 & \frac{-1}{L}+\frac{L}{6} & \frac{1}{L}+\frac{L}{3} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{A}^3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{L}+\frac{L}{3} & \frac{-1}{L}+\frac{L}{6} \\ 0 & 0 & \frac{-1}{L}+\frac{L}{6} & \frac{1}{L}+\frac{L}{3} \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{L}+\frac{L}{3} & \frac{-1}{L}+\frac{L}{6} & 0 & 0 \\ \frac{-1}{L}+\frac{L}{6} & 2\left(\frac{1}{L}+\frac{L}{3}\right) & \frac{-1}{L}+\frac{L}{6} & 0 \\ 0 & \frac{-1}{L}+\frac{L}{6} & 2\left(\frac{1}{L}+\frac{L}{3}\right) & \frac{-1}{L}+\frac{L}{6} \\ 0 & 0 & \frac{-1}{L}+\frac{L}{6} & \frac{1}{L}+\frac{L}{3} \end{bmatrix}$$

From Eq. (7.78) we have:

$$\boldsymbol{A}\begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix} = \begin{bmatrix} -\frac{d\hat{h}}{dx}\Big|_{x=0} \\ 0 \\ 0 \\ -\frac{d\hat{h}}{dx}\Big|_{x=3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

h_1 and h_4 are known; therefore, from the above equation set we have:

$$2\left(\frac{1}{L}+\frac{L}{3}\right)h_2+\left(-\frac{1}{L}+\frac{L}{6}\right)h_3=0$$

$$\left(-\frac{1}{L}+\frac{L}{6}\right)h_2+2\left(\frac{1}{L}+\frac{L}{3}\right)h_3=-10\left(-\frac{1}{L}+\frac{L}{6}\right)$$

If $L = 1$, the results are as follows:

$$h_2 = 1.08 \text{ m and } h_3 = 3.46 \text{ m}$$

7.4.2.3 Grid Design

The finest mesh spacing and time steps should be selected to reduce numerical errors, while computational time and limitations on computer memory should also be considered. In rectangular grid design, the length-to-width ratio (aspect ratio) of cells or elements should be kept as close as possible to one in order to prevent numerical instability or errors (Delleur, 1999). In the case of a triangular finite element, the angles should be less than 22.5° (Torak, 1993). To decrease the computational cost, variably spaced mesh is usually used with a fine grid in the following areas:

- Where the greatest accuracy is desirable
- Where the gradients are steepest (stress points)
- Where the data are concentrated

It is generally advisable for the increasing factor for mesh spacing in two adjacent elements or cells to be less than two (Delleur, 1999). Similarly, in transient simulation different time steps can be used to optimize computational resources.

7.4.2.4 Parameter Estimation and Model Calibration

Groundwater simulation models require considerable data to define all nodal parameters. The accuracy of simulation depends on the reliability of the estimated parameters as well as accuracy of model and boundary conditions. As the parameters of groundwater models cannot be measured directly, they can be estimated from historical data using an inverse parameter estimation procedure (ReVell and McGarity, 1997).

In deterministic groundwater simulation models, the objective of the calibration procedure is to minimize the difference between observed aquifer response data (such as water table variation) and corresponding values calculated by the model. The comparison between observed and calculated values is usually subjective and it should be considered that a good match does not necessarily reflect the validity of the model.

Model calibration is usually a time-consuming, trial-and-error procedure that modifies the model output by changing the aquifer recharges and discharges, aquifer properties, and initial and boundary conditions. In this case, the uncertainties in sources, sinks, and initial and boundary conditions, as well as the uncertainties in aquifer properties, should be evaluated and considered in the calibration procedure. Therefore, the experience and engineering judgment of the modeler are important factors in efficient calibration of the model.

The efficiency of model calibration can be improved by automated parameter estimation techniques that use the least-square deviation as a criterion to obtain the estimate of system parameters.

Yeh (1992) classified the solution algorithms of parameter identification in groundwater systems as being part of either the *output error criterion method* or *the equation error criterion method*. In the equation error criterion method, finite-element or finite-difference approximation of the groundwater equation is used to obtain a set of algebraic equations in terms of unknown parameters. The unknown parameters are determined by minimizing the sum of squares of errors for the equation, using the observed or interpolated data at each node of the domain (ReVelle and McGarithy, 1997). The output error criterion method is an optimization problem that minimizes a given output error criterion during identification of the flow model parameters.

7.4.2.5 Model Verification

The calibrated model should be verified before using it for groundwater simulation. In this case, the observed data that have not been used in the model calibration process are compared with the results of simulation model to evaluate the reliability and accuracy of model prediction. As the observed data from groundwater systems are usually limited, the model verification process usually suffers from data deficiency. In this case, the unverified model should be applied only if the sensitivity analyses for both model calibration and prediction are performed (ReVelle and McGarity, 1997).

7.4.2.6 Predictions and Post Audits

The reliability of a calibrated model can be evaluated by the accuracy of the model predictions. Future stresses and their uncertainties are the major sources of prediction error. For example, boundary conditions and processes that are insignificant under current and past stress regimes may become important when existing stresses are increased or new stresses are added. Therefore, if a model is continuously used for prediction of future conditions of a groundwater system, the model should be periodically audited or calibrated using the field monitoring data (Delleur, 1999). In this way, the model can incorporate new information such as changes in assumed groundwater stresses.

7.5 MANAGEMENT TOOLS

7.5.1 Groundwater Simulation Models

7.5.1.1 Graphic Groundwater

Graphic Groundwater is a three-dimensional groundwater model written in the C++ programming language. This is a simplified version of the U.S. Geological Survey Three-Dimensional Modular Finite Difference Groundwater Flow Model (MODFLOW; see McDonald and Harbaugh, 1988), and enhanced data input and display features have been added (Esling et al., 1993). This program is designed to simplify model development and data input for groundwater modeling to develop maps and diagrams within Graphic Groundwater and to provide graphics and text files that can be processed outside of Graphic Groundwater.

The first step toward creating a model is the development of a grid system, then layers are added and parameters of hydraulic conductivity, initial head, storage coefficient, top elevation, and bottom elevation are specified for each cell. The number of stress periods over which the model will run and the time of their operation should be set.

Results could be analyzed through contour plots, if desired. Superimposing a grid over a graphic image or digitized line drawing is one of the most powerful features of Graphic Groundwater. These images are generally two-dimensional plan views representing the area that is going to be modeled.

Graphic Groundwater solves the same system of equations as MODFLOW, but the discretization convention is slightly different. Graphic Groundwater requires the top and bottom elevation of a cell in both confined and unconfined aquifers. It calculates cell thickness from this information. The thickness is not implied in various hydrologic parameters, thus Graphic Groundwater accommodates only the vertical discretization scheme between two end members such that cell boundaries are coincident with hydrologic boundaries.

7.5.1.2 MODFLOW

MODFLOW is a three-dimensional, finite-difference groundwater model that was first released in 1984. It has a modular structure that allows the user to modify it and adapt the code for a particular application. Many new capabilities have been added to the original model in recent versions. MODFLOW-2000 simulates steady and unsteady flow in an irregularly shaped flow system in which aquifer layers can be confined, unconfined, or a combination of confined and unconfined. Flow from external stresses, such as flow to wells, surface recharge, evapotranspiration, flow to drains, and flow through river beds, can be simulated. Hydraulic conductivities or transmissivities for any layer may differ spatially and be anisotropic (restricted to having the principal directions aligned with the grid axes), and the storage coefficient may be heterogeneous. Specified head and flux boundaries can be simulated by this model. It is also capable of modeling head-dependent flux across the

outer boundary of the model, which allows water to be supplied to a boundary block in the modeled area at a rate proportional to the current head difference between a source of water outside the modeled area and the boundary block.

In addition to simulating groundwater flow, the scope of MODFLOW-2000 has been expanded to incorporate related capabilities such as solute transport and parameter estimation. In this model, the groundwater flow equation is solved using the finite-difference approximation. The flow region is subdivided into blocks in which the medium properties are assumed to be uniform. In plan view, the blocks are made from a grid of mutually perpendicular lines that may be variably spaced. Model layers can have varying thickness. A flow equation is written for each block, or *cell*. Several methods are provided for solving the resulting matrix problem; the user can choose the best one for a particular problem. Flow rate and cumulative volume, which are balanced from each type of inflow and outflow, are computed for each time step.

7.5.2 Safe Yield Estimation

In groundwater management, one of the main goals is to evaluate the maximum annual groundwater yield that can be withdrawn without producing undesirable effects. The concept of *safe yield* was first defined by Lee (1915) as "the limit to the quantity of water which can be withdrawn regularly and permanently without depletion of storage reserves." Early in this century, "undesirable effects" meant violation of water laws or adverse effects on the supplies of adjacent landowners using the same aquifer due to overdraft of the aquifer, but today undesirable effects include any adverse effects on ecosystems. The simple concern of overdraft can be approached using the hydrologic equilibrium equation as follows:

$$I = O - \Delta S \tag{7.83}$$

where I is inflow, O is the outflow, and ΔS is the change in storage. Inflow is the summation of surface and subsurface inflow, precipitation, and imported water. Outflow is the summation of surface and subsurface outflow, consumptive use, and exported water (Todd, 1980).

To determine the safe yield in a basin, calculation of subsurface inflow and outflow is usually the most difficult task because it is impossible to measure them directly. Another complicating factor is that the definitions of adverse effects are subjective and subject to temporal and spatial interpretations. Acceptable definitions of adverse effects can change in the future due to changes in demographics, land use, and political and social conditions.

7.5.3 Optimization Models for Groundwater Management

The optimal water supply to municipal, industrial, and agricultural water demands considering the physical, socioeconomic, and environmental constraints are the

major objective of groundwater planning and management. The following planning problems are associated with groundwater supply management (Yeh, 1992):

- Determination of an optimal pumping pattern, such as location of pumping wells and their pumping rate, to satisfy water demands
- Timing and staging of well system development (capacity expansion) with respect to future water demands
- Design of water transfer facilities for optimal allocation and distribution of water to demand points in the basin

This section presents some typical optimization models for groundwater operation management, capacity expansion, and joint use of surface and groundwater resources.

7.5.3.1 Optimization Model for Groundwater Operation

A groundwater operation model can provide the optimal policies for groundwater operation such as optimal extraction, allocation, and optimal recharge of the aquifer considering physical, environmental, and socioeconomic constraints. An objective function of the optimization model could be to minimize the total discounted operational costs as follows:

$$\text{Min } Z = \sum_{t=1}^{T} \left(\frac{1}{1+\alpha} \right)^{t} \sum_{w \in W} C_w \left(Q_w^t, h_w^t \right) \tag{7.84}$$

where:

T is the number of time steps in the planning horizon.
α is the interest rate.
W is the set of well sites.
$C_w(Q_w^t, h_w^t)$ is the water extraction cost from well site w and is a nonlinear function of discharge Q_w^t and head h_w^t at each time step t.

This objective function should be minimized taking into consideration groundwater response equations and constraints of the system.

The response equations for each planning period in a linear, distributed parameter groundwater system can be expressed as follows (Willis and Yeh, 1987):

$$\boldsymbol{h}^t - A_1 \boldsymbol{h}^{t-1} - A_2 \boldsymbol{g}(\boldsymbol{Q}^{t-1}) = 0 \qquad \forall t \tag{7.85}$$

where $\boldsymbol{h}^t$ and $\boldsymbol{h}^{t-1}$ are vectors of groundwater head at time steps t and $t-1$, A_1 and A_2 are coefficient matrices, and vector $\mathbf{Q}^{t-1}$ is defined as follows:

$$\boldsymbol{Q}^{t-1} = (Q_1^{t-1}, Q_2^{t-1}, ..., Q_w^{t-1})^T \qquad \forall t, w \in W \tag{7.86}$$

The following constraints can be considered in a groundwater operation optimization model:

- *Supplying demands*

$$\sum_{w \in W} Q_w^t \geq D^t \qquad \forall t \tag{7.87}$$

where D^t is the total water demand at time step t.

- *Discharge capacity for each well*

$$0 \leq Q_w^t \leq Q_w^{\max} \qquad \forall t, w \in W \tag{7.88}$$

where $Q_w^{\max}$ is the discharge capacity of well w.

- *Lower bound constraints on the head levels*

$$h_w^t \geq h_w^{\min} \qquad \forall t, w \in W \tag{7.89}$$

where $h_w^{\min}$ is the minimum groundwater head at well site w. These constraints are considered to ensure that excessive depletion of the aquifer does not occur.

- *Non-negativity of the decision variables*

$$\boldsymbol{h}^t, Q_w^t \geq 0 \qquad \forall t, w \in W \tag{7.90}$$

The above planning model is the basic framework for groundwater systems planning and management.

7.5.3.2 Optimization Model for Capacity Expansion

Capacity expansion and the timing and staging of well field development are typical problems in groundwater systems planning. In the problems presented here, it is assumed that the well sites have been developed and operation policies are developed for existing conditions; therefore, capital cost can be neglected when developing the operation policies and should be considered in capacity expansion models. The objective function of the capacity expansion model — minimization of total discounted capital and operational cost — is as follows:

$$\text{Min } Z = \sum_{t=1}^{T}\left(\frac{1}{1+\alpha}\right)^{t}\sum_{w\in W}\left(KC(Q_w^{\max})X_w^t + C_w\left(Q_w^t, h_w^t\right)Y_w^t\right) \tag{7.91}$$

where:

T is the number of time steps in the planning horizon.
α is the interest rate.
W is the set of well sites.
$KC(Q_w^{\max})$ is the capital cost of well site w, which is considered as a function of well capacity ($Q_w^{\max}$) and $C_w(Q_w^t, h_w^t)$ is the water extraction costs from well site w, which is a nonlinear function of discharge Q_w^t and head h_w^t at each time step t.
X_w^t and Y_w^t are 0/1 integer variables. X_w^t indicates whether or not well site w is developed in period t. X_w^t is equal to 1 if well site w is developed in time period t and equal to 0 otherwise. Similarly, Y_w^t is equal to 1 if well site w is in operation and equal to 0 otherwise.

This objective function should be minimized taking into consideration groundwater response equations and constraints of the system:

- *Response equations for each planning period*

$$\boldsymbol{h}^t - A_1\boldsymbol{h}^{t-1} - A_2\boldsymbol{g}(\boldsymbol{Q}^{t-1}) = 0 \qquad \forall t \tag{7.92}$$

- *Supplying water demands*

$$\sum_{w\in W} Y_w^t Q_w^t \geq D^t \qquad \forall t \tag{7.93}$$

where D^t is the total water demand in time step t.

- *Constraint of well capacity for each well site w*

$$0 \leq Q_w^t \leq Q_w^{\max} \qquad \forall t, w \in W \tag{7.94}$$

where $Q_w^{\max}$ is the discharge capacity of well w.

- *Lower bound constraints on head levels*

$$h_w^t \geq h_w^{\min} \qquad \forall t, w \in W \tag{7.95}$$

where $h_w^{\min}$ is the minimum groundwater head at well site w. These constraints are considered to ensure that excessive depletion of the aquifer does not occur.

- *0/1 integer variables*

$$X_w^t, Y_w^t = [0,1] \tag{7.96}$$

- The maximum number of developed wells over the planning horizon is assumed to be 1; therefore,

$$\sum_{t=1}^{T} X_w^t \le 1 \qquad w \in W \tag{7.97}$$

- *Limitations on total capital and operational costs at each planning period*

$$\sum_{w \in W} \left(KC(Q_w^{\max}) X_w^t + C_w\left(Q_w^t, h_w^t\right) Y_w^t \right) \le C_{total,t} \qquad \forall t \tag{7.98}$$

where $C_{total,t}$ is the total capital and operational cost in planning period t.

- *Non-negativity of the decision variables*

$$\boldsymbol{h}^t, Q_w^t \ge 0 \qquad \forall t, w \in W \tag{7.99}$$

Because of uncertainties regarding future operational and capital cost functions and projected water demands, the developed optimal policies should be revised when more information becomes available.

7.5.3.3 Optimization Model for Water Allocation

The groundwater allocation problem provides the optimal water distribution for various demands such as agricultural, industrial, and municipal, in a river basin with minimal effect on the environment and water users. The objective function of this model maximizes the net discounted benefit of the water supply as follows:

$$\text{Max} \quad Z = \sum_{t=1}^{T} \left(\frac{1}{1+\alpha} \right)^t \sum_i \left\{ \int_0^{Q_i^t} D_i\left(Q_i^t\right) dQ - \sum_{w \in W} C_w\left(Q_w^t, h_w^t\right) \right\} \tag{7.100}$$

where:

T is the number of time steps in the planning horizon.
α is the interest rate.
W is the set of well sites.
$D_i(Q_i^t)$ is the groundwater aggregate demand function of subregion i, which depends on Q_i^t. The area beneath the demand function represents willingness to pay for a specific quantity of water (Q_i^t). In other words, this function relates the price and quantity demanded of the water supply (Willis and Yeh, 1987).
Q_i^t is the water pumped from subregion i.
$C_w(Q_w^t, h_w^t)$ is the water extraction cost from well site w, which is a nonlinear function of discharge Q_w^t and head h_w^t at each time step t.

This objective function should be maximized considering the groundwater response equations and constraints of the system:

- *Response equations for each planning period*

$$\boldsymbol{h}^t - A_1\boldsymbol{h}^{t-1} - A_2\boldsymbol{g}(\boldsymbol{Q}^{t-1}) = 0 \qquad \forall t \tag{7.101}$$

- *Constraint of water balance in each subregion*

$$Q_i^t = \sum_{w \in W} Q_w^t \tag{7.102}$$

- *Constraint of well capacity for each well site w*

$$0 \le Q_w^t \le Q_w^{\max} \qquad \forall t, w \in W \tag{7.103}$$

where $Q_w^{\max}$ is the discharge capacity of well w.

- *Non-negativity of the decision variables*

$$\boldsymbol{h}^t, Q_w^t, Q_i^t \ge 0 \qquad \forall t, w \in W, \forall i \tag{7.104}$$

The water allocation model can provide a schedule for optimal groundwater extraction.

7.5.3.4 Optimization Model for Conjunctive Water Use

The conjunctive water use model provides an optimal allocation schedule for ground and surface water resources in a region to meet different water demands. The objective function maximizes the net discounted economic benefit as follows:

$$\text{Max} \quad Z = \sum_{t=1}^{T} \left(\frac{1}{1+\alpha} \right)^{t} \left\{ \sum_{l} \left[\int_{0}^{Q_l^t} D_l\left(Q_l^t\right) dQ - \sum_{m} g_{ml}^t\left(G_{ml}^t\right) - \sum_{n} f_{nl}^t\left(S_{nl}^t\right) \right] \right\} \tag{7.105}$$

where:

l, m, n are indexes that define different demands, groundwater resources, and surface water resources, respectively, in the region.

$D_l(Q_l^t)$ is the demand function, which depends on the water allocated to demand l (Q_l^t).

Q_l^t and $g_{ml}^t(G_{ml}^t)$ is the cost function of groundwater allocation from resource m to demand l, which depends on the amount of groundwater allocated to that demand, in period t, (G_{ml}^t).

$f_{nl}^t(S_{nl}^t)$ is the cost function of surface water allocation from resource n to demand l, which depends on the amount of surface water allocated to that demand, in period t, (S_{nl}^t).

The model is constrained by the response equations and physical limitations:

- *Response equations for each groundwater resource*

$$\boldsymbol{h}_{\boldsymbol{m}}^{\mathrm{t}} - A_1^m \boldsymbol{h}_{\boldsymbol{m}}^{\mathrm{t\text{-}1}} - A_2^m \boldsymbol{g}\left(\sum_{m} G_{ml}^{t-1} \right) = 0 \tag{7.106}$$

- *Surface water balance equation,* assuming no hydraulic interaction between the surface and groundwater systems:

$$\overline{S}_n^t = \overline{S}_n^{t-1} + R_n^t - \sum_{l} S_{nl}^t \tag{7.107}$$

where $\overline{S}_n^t$ is the volume of water in surface storage n at the end of time period t and R_n^t is the inflow to surface storage n during period t.

- *The balance equations relating the groundwater allocations and pumping schedules*

$$\sum_{l} G_{ml} = \sum_{w \in W} Q_{w,m}^t \qquad , \forall m \tag{7.108}$$

where $Q_{w,m}^t$ is the pumping rate from well w in groundwater resource m, and W defines the well sites of the region.

- *Well capacity limitation*

$$Q_{w,m}^{t} \le Q_{w,m}^{\max} \qquad ,\forall m, \forall t, w \in W \tag{7.109}$$

- *Lower bound constraints on the head levels*

$$h_{w}^{t} \ge h_{w}^{\min} \qquad \forall t, w \in W \tag{7.110}$$

  These constraints are considered to ensure that excessive depletion of the aquifer does not occur.

- *Non-negativity of the decision variables*

$$\boldsymbol{h}_{m}^{t}, Q_{w,m}^{t}, Q_{l}^{t}, G_{ml}^{t}, S_{nl}^{t} \ge 0 \qquad w \in W, \forall l, m, n, t \tag{7.111}$$

This optimization problem can be solved using different methods such as nonlinear or dynamic programming.

Example 7.9

Determine the average monthly allocation of groundwater and surface water to four agricultural zones shown in Figure 7.15. As can be seen in this figure, a channel can transfer surface water to zone 4. The water demands of zone 3 cannot be supplied from the rivers. The monthly discharge of rivers and the volume of recharge by precipitation during a 2-year period and the gross water demands of the agricultural zones are presented in Tables 7.2 and 7.3, respectively. The equations of the average groundwater table fluctuations, based on the discharge from agricultural wells and recharge from precipitation, are presented in Table 7.4. In this table, $G_t(i)$ is the groundwater discharge from agricultural zone i in period t, $Q_t(1)$ and $Q_t(2)$ are the allocated water to agricultural zones 1 and 2 in period t, $Q_t(3)$ is the surface water transferred to zone (4) in period t, and P is the groundwater recharge from precipitation (million m^3). The effects of other parameters such as underground inflow and outflow on the fluctuations of groundwater table are considered to be negligible. The acceptable (without any cost) range for cumulative groundwater table variation in each zone is ± 4 meters. The maximum capacity of water transfer channel is 4 m^3/sec, and it is assumed that 90 percent of transferred water is supplied by River 1. The initial groundwater table depths in zones 1, 2, 3, and 4 are considered to be 17, 8.1, 12.6, and 67 m, respectively. The objectives of the developed model are supplying water to agricultural lands, minimizing pumping costs, and controlling average groundwater table fluctuations in agricultural zones. Calculate the optimal allocated water to agricultural zone, from surface and groundwater resources.

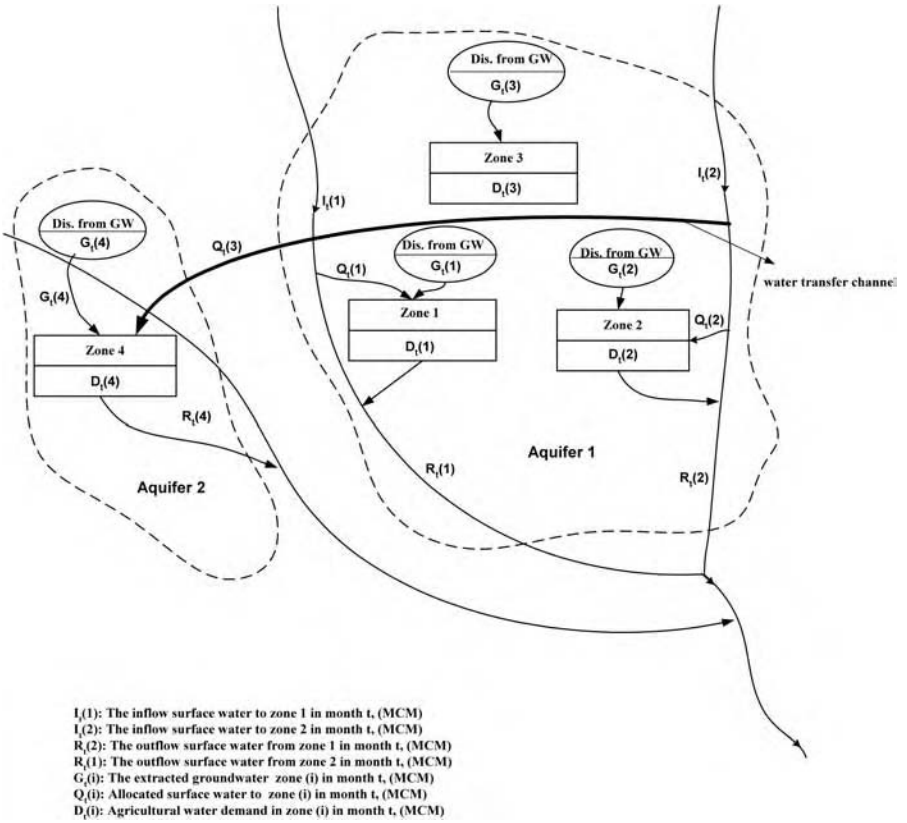


FIGURE 7.15 Components of surface and groundwater resources for Example 7.9.

Solution: The pumping cost can be considered as a linear function of pumping power (P_{pump}) as follows:

$$P_{pump} = \frac{G.H}{102\eta} \tag{7.112}$$

where P_{pump} is the pump power (W), G is the pumping discharge (m^3/sec), H is the groundwater table depth (m), and η is the pump efficiency.

Because the objective function and the constraints are nonlinear, the dynamic programming (DP) method can be used effectively in this study. The recursive function of the DP model is developed as follows:

$$\begin{aligned} f_t(G_t(i), Q_t(i)) = Min\Big\{ & C_t(G_t(i), Q_t(i), H_{t-1}(i), I_t, Q_t) \\ & + f_{t-1}^{\,*}(G_{t-1}(i), Q_{t-1}(i), H_{t-2}(i), I_{t-2}, Q_{t-1}) \Big\} \end{aligned} \tag{7.113}$$

TABLE 7.2
Values of Monthly River Discharge and Precipitation

Water Year	Month	Discharge of River 1 (million m^3)	Discharge of River 2 (million m^3)	Recharge of Aquifer 1 by Precipitation (million m^3)	Recharge of Aquifer 2 by Precipitation (million m^3)
1	October	13.48	15.49	0.7	0.05
1	November	6.16	16.51	1.2	0.2
1	December	9.38	22.78	2	0.18
1	January	14.34	20.19	2.5	0.19
1	February	15.9	27.18	3	0.3
1	March	18.17	26.25	3	0.31
1	April	10.08	26.99	2.5	0.3
1	May	7.11	26.16	2	0.15
1	June	5.3	23.14	0.5	0.04
1	July	3.76	20.29	0	0.01
1	August	3.23	19.65	0	0.2
1	September	2.85	17.43	0.5	0.05
2	October	18.83	30.18	0.6	0.06
2	November	7.44	26.23	1.2	0.2
2	December	8.58	26.64	2.2	2.2
2	January	10.89	26.45	2.3	2.3
2	February	14.72	25.61	3.2	3.2
2	March	25.79	20.22	2.9	2.9
2	April	35.17	48.29	2.6	2.6
2	May	15.07	35.09	2	2
2	June	9.47	28.07	0.56	0.56
2	July	19.08	23.16	0.2	0.2
2	August	37.51	19.16	0	0
2	September	30.25	16.47	0.51	0.51

TABLE 7.3
Monthly Agricultural Water Demands and Downstream Water Rights

Month	Zone 1 (million m³)	Zone 2 (million m³)	Zone 3 (million m³)	Zone 4 (million m³)	Downstream Water Rights (million m³)
October	8.536	10.07	0.64	3.01	1.86
November	8.494	8.59	0.48	3.82	3.43
December	3.45	2.57	0.2	1.41	1.045
January	2.448	2.03	0.153	1.24	0.965
February	2.371	1.79	0.133	0.81	0.37
March	4.5	3.33	0.353	1.32	0.54
April	11.44	9.75	1.176	6.15	6.92
May	34.64	31.76	3.44	20.58	28.57
June	36.6	35.22	4.02	19.12	27.6
July	19.45	18.83	2.53	4.93	8.92
August	23.9	25.86	2.66	4.97	5.78
September	21.31	23.49	1.68	4.26	3.76

TABLE 7.4
Equations of Average Groundwater Table Fluctuation

Zone 1	$\Delta L_1 = 1.9\times10^{-4}G(1)+1.2e^{-3}G(2)-2e^{-5}G(3)^2+7.8e^{-4}G(3)-0.02P-0.03$
Zone 2	$\Delta L_2 = 1\times10^{-6}G(1)+3.7e^{-3}G(2)-4e^{-5}G(3)^2+1.9e^{-4}G(3)-0.018P-0.07$
Zone 3	$\Delta L_3 = 1\times10^{-7}G(1)+1.1e^{-3}G(2)-4.1e^{-4}G(3)-0.01P-0.02$
Zone 4	$\Delta L_4 = -0.0149G(4)^2+0.41G(4)-7e^{-5}Q(3)^2-0.028Q(3)-0.12$

Note: Positive values represent the groundwater table drawdown.

where:

i is the index of zone $i = 1, \ldots, 4$.
$G_t(i)$ is the amount of groundwater extracted from zone i in month t (million m³).
$Q_t(i)$ is the surface water allocated to zone i in month t (million m³).
$C_t(G_t(i), Q_t(i))$ is the cost of operation during time period t.
$f^*_{t-1}(G_{t-1}(i), Q_{t-1}(i))$ is the minimum total operational cost until the end of time period $t-1$.
$H_t(i)$ is the depth of groundwater table in agricultural zone i at the end of period t.
I_t/O_t is the total recharge/discharge of aquifers during the period t.

The cost of operation in each period is estimated as:

$$C_t(G_t(i), Q_t(i), H_{t-1}(i), I_t, O_t) = \sum_{i=1}^{4} Loss_t(i) \tag{7.114}$$

$$Loss_t(i) = \alpha(D_t(i) - Q_t(i) - G_t(i))^2 + \beta(G_t(i) \cdot H_t(i)) + \gamma\left(\left|L_t(i)\right| - L_{\max}(i)\right)^2 \tag{7.115}$$

$$\alpha = 0 \qquad \text{if} \quad D_t(i) \le (Q_t(i) + G_t(i)) \tag{7.116}$$

$$\gamma = 0 \qquad \text{if} \quad \left|L_t(i)\right| \le L_{\max}(i) \tag{7.117}$$

where:

$D_t(i)$ is the agricultural water demand in zone i and time period t.
$L_t(i)$ is the total variation of the water table level until the end of time period t.
$L_{max}(i)$ is the maximum allowable cumulative groundwater table fluctuation in agricultural zone i.
$H_t(i)$ is the depth of the groundwater table in agricultural zone i at the end of month t.
α, β, γ are the weights of the objectives (constant).

The constraints of the model are as follows:

- *Control of groundwater level fluctuations*

$$H_t(i) = H_0(i) + \sum_{t=1}^{T} \Delta L_t(i) \qquad i = 1, ..., 4 \tag{7.118}$$

$$L_t(i) = \sum_{t=1}^{T} \Delta L_t(i) \qquad i = 1, ..., 4 \tag{7.119}$$

$$\Delta L_t(i) = F_i(G_t(i), Q_t(i), I_t, O_t) \tag{7.120}$$

where:

$H_0(i)$ is initial depth of groundwater table in agricultural zone i.
$\Delta L_t(i)$ is the variation in water table level for time period t in agricultural zone i (drawdown is considered to be positive).
$F_t(G_t(i), Q_t(i), I_t, O_t)$ is groundwater table fluctuations in agricultural zone i for time period t and are functions of total discharge (O_t) and total recharge (I_t) of the aquifers.

- *Estimation of surface water outflow from agricultural zones 1 and 2*

$$R_t(i) = I_t(i) - Q_t(i) - (1 - \frac{\psi}{100})Q_t(3) \qquad i = 1, 2 \tag{7.121}$$

where ψ is the percent of transferred water to zone 4 from surface water resources of zone 2. $I_t(i)$ is the surface water inflow to zone i for time period t (million m^3).

- *Minimum instream flow requirements downstream of agricultural zones*

$$\sum_{i=1}^{2} R_t(i) \geq R_{t,\min} \tag{7.122}$$

where $R_{t,min}$ is the minimum instream flow required in downstream in time period t, and $R_t(i)$ is the surface water outflow from zone i for time period t (million m^3).

- *Maximum capacity of canal*

$$Q_t(3) \leq Q_{t,\max} \tag{7.123}$$

Because the problem has multiple objectives, the coefficients α, β, and γ are the relative weights of objectives which should be determined in order to make these objectives comparable. The optimal average water allocated to agricultural zones in July (a dry month) for $\alpha = 1$, $\beta = 0.005$, and $\gamma = 120$ and the corresponding cumulative groundwater table variations are presented in Table 7.5 and Figure 7.16 (for more details, see Karamouz et al., 2002).

7.6 CONFLICT ISSUES IN GROUNDWATER SYSTEMS

The operation of groundwater systems is a multiple-objective problem, and some economic, hydraulic, water quality, or environmental objectives are usually in conflict. For example:

- Several demand points exist, and the water supplied to one of these demand points cannot be used by the others; therefore, a major conflict issue in groundwater operation occurs when the groundwater storage is not capable of supplying all of the demands.
- Extra discharge of the aquifer or high variation of the groundwater table can cause problems, such as settlement of buildings and the ground surface.
- Recharge of the aquifer with polluted water, such as occurs due to infiltration of agricultural returnflows, disposal of sewage or water treatment plant sludge, and sanitary landfills can produce environmental problems.

TABLE 7.5
Average Surface and Groundwater Allocated in July

Source	Zone 1 (million m^3)	Zone 2 (million m^3)	Zone 3 (million m^3)	Zone 4 (million m^3)
Surface water	5.8	15.6	—	—
Groundwater	21.2	3.0	3.0	1.0
Transferred water	—	—	—	3

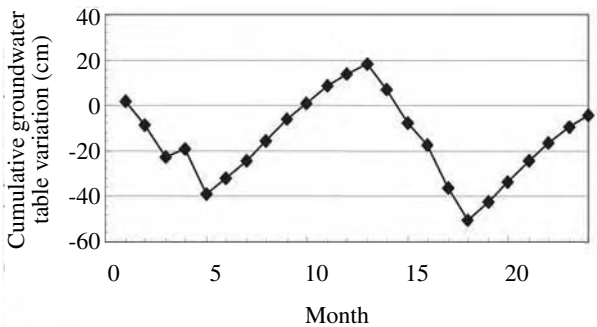


FIGURE 7.16 Cumulative groundwater table variation in zone 1 for Example 7.9.

The optimal operation policies of groundwater systems can be developed using conflict resolution models considering groundwater system hydraulic or water quality response equations and possible well capacity, hydraulic gradient, and water demand requirements.

A common conflict issue in groundwater systems planning and operation occurs when the aquifer should supply water to different demand points for different purposes with constraints on groundwater table variations and groundwater quality. Consider an aquifer system that supplies the following demands:

- Domestic water demand
- Agricultural water demand
- Industrial water demand

Even though water resources planning for operation of wells is done by a specific agency on a regional basis (called here the Department of Water Supply), the water allocation schemes defined and imposed by this department are greatly influenced by other agencies, such as these generic examples:

- Department of Agriculture
- Department of Industries
- Department of Environmental Protection

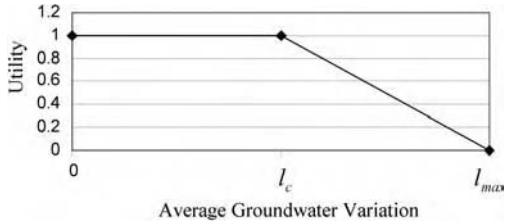


FIGURE 7.17 Utility function for the range of average groundwater variation.

The utility function for each of these agencies should be considered when formulating the conflict resolution problems; therefore, all of the conflict issues and the responsible agencies should be recognized in the first step of conflict resolution studies.

The next step is to define the utility function of each agency. For example, the Department of Water Supply could define the utility function for a range of average groundwater table variations, as shown in Figure 7.17, where the most favorable range of average groundwater table variation is from 0 to l_c m. This range can be estimated based on the hydrogeologic conditions of an aquifer, as well as the estimates of future demands. The decreasing segment on the right-hand side of this graph also shows that the utility of the Department of Water Supply will decrease when the groundwater table variation is increased.

All of the agencies selected in the first step should provide their favorable range of water supply for various demands, groundwater quality, or water table fluctuation. Besides the different priorities of these agencies, each agency has a specific level of authority with respect to changing the water allocation schemes within the political and institutional climate of each region. The final step is to formulate the conflict resolution problem. For this purpose, various methods can be used, which are briefly explained in Chapter 2.

Example 7.10

In an unconfined aquifer system, the following agencies are affected by the decision to discharge from an aquifer to supply water demands:

- Agency 1: Department of Water Supply
- Agency 2: Department of Agriculture
- Agency 3: Department of Industries
- Agency 4: Department of Environmental Protection
- Agency 5: Department of Domestic Water Use

The Department of Water Supply has a twofold role — namely, to allocate water to different purposes and control the groundwater table variations. The decision makers in these agencies are asked to set their utility functions, which are shown in Figures 7.18 to 7.20. The analyst should apply the weights presented in Table 7.6 to consider the roles and levels of authority of the agencies in the political climate of that region.

The initial surface area, volume, and total dissolved solids (TDS) concentration for the water in the aquifer are 600 km^2, 1440 million m^3, and 1250 mg/L, respectively. The net underground inflow is 1100 million m^3/yr, with a TDS concentration equal to 1250 mg/l. It is assumed that 60% of the allocated water returns to the aquifer as return flow and the average TDS concentration of the return flow is 2000 mg/L. The average storage coefficient of the aquifer is 0.06. Find the most appropriate water allocation scheme for this year using Nash's bargaining theory.

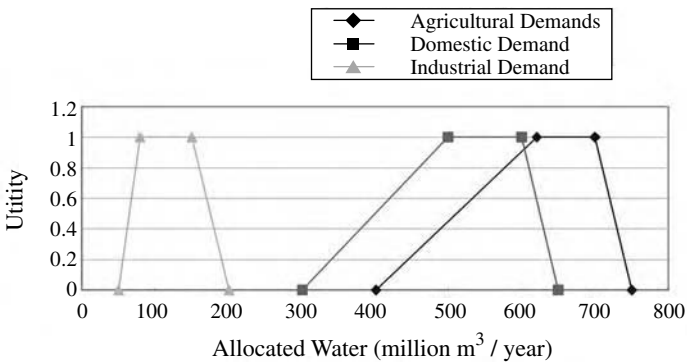


FIGURE 7.18 Utility function of different agencies for the range of water allocated to different demands.

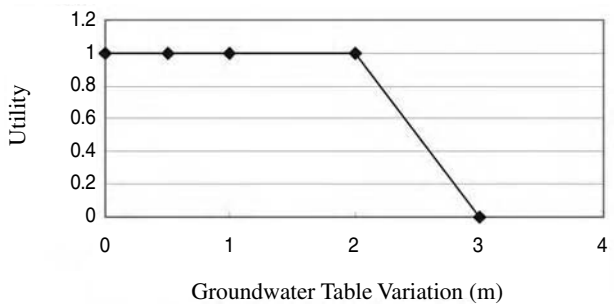


FIGURE 7.19 Utility function of Agency 1 for the range of average groundwater variation.

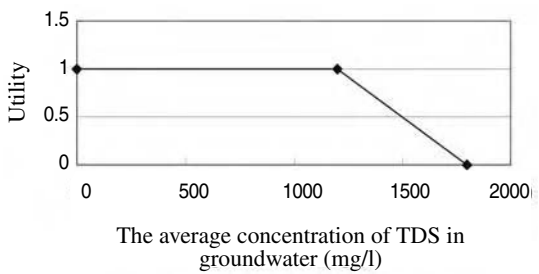


FIGURE 7.20 Utility function of Agency 4 for the average concentration of TDS in the groundwater.

TABLE 7.6
Relative Weights or Relative Authority of Agencies

Agencies	Relative Weight Case 1	Relative Weight Case 2
Department of Agriculture	0.133	0.17
Department of Domestic Water Use	0.33	0.2
Department of Water Supply	0.2	0.23
Department of Environmental Protection	0.2	0.3
Department of Industries	0.133	0.1

Solution: The nonsymmetric Nash solution of problem is the unique optimal solution of the following problem:

$$\prod_{i=1}^{5} (f_i - d_i)^{w_i} \tag{7.124}$$

$$d_i \le f_i \le f_i^* \quad \forall i \tag{7.125}$$

where W_i is the relative weight, f_i is the utility function, d_i is the disagreement point, and f_i^* is the ideal point of player (agency) i. This objective function should be maximized considering the constraints of groundwater surface variation and groundwater quality that are as follows:

1. The average groundwater table variation:

$$\Delta l \approx (q_1 + q_2 + q_3 - IN) \times (1/S) \times (1/A) \tag{7.126}$$

where Δl is the average groundwater table variation (m) during the planning year, q_1, q_2, q_3 are annual agricultural, domestic, and industrial groundwater withdrawal (million m^3), IN is net annual underground inflow (million m^3), S is storage coefficient, and A is the area of aquifer (km^2). Therefore,

$$\Delta l \approx \frac{(q_1 + q_2 + q_3 - 1100)}{0.06 \times 600}$$

2. The average groundwater quality:
Considering the TDS as the indicator water quality variable, the average TDS concentration of groundwater at the end of the year can be simulated as follows:

$$C_{new} = \frac{(V_{in.} \times C_{in.} + IN \times C_{IN} + \alpha_1 \times C_1 \times q_1 + \alpha_2 \times C_2 \times q_2 + \alpha_3 \times C_3 \times q_3 - C_{mean}(q_1 + q_2 + q_3))}{(V_{in.} + IN + (\alpha_1 - 1) \times q_1 + (\alpha_2 - 1) \times q_2 + (\alpha_3 - 1) \times q_3)} \tag{7.127}$$

and

$$C_{mean} = (C_{in.} + C_{new}) / 2 \tag{7.128}$$

where:

$V_{in.}$ and $C_{in.}$ are the initial volume and TDS concentration of the groundwater, respectively.
C_{in} is the TDS concentration of net groundwater inflow.
α_1, α_2, α_3 are the return flows (%) of annual agricultural, domestic, and industrial groundwater withdrawals, respectively.
C_1, C_2, C_3 are the average TDS concentrations of agricultural, domestic, and industrial groundwater withdrawal, respectively.
C_{mean} is the average TDS concentration of groundwater during the planning year.
C_{new} is the TDS concentration of the aquifer at the end of the year.

Therefore,

$$C_{new} = \frac{(1440 \times 1250 + 1100 \times 1250 + 0.6 \times 2000 \times q_1 + 0.6 \times 2000 \times q_2 + 0.6 \times 2000 \times q_3 - C_{mean}(q_1 + q_2 + q_3))}{(1440 + 1100 + (0.6 - 1)(q_1 + q_2 + q_3)}$$

and

$$C_{mean} = (1250 + C_{new}) / 2$$

This nonlinear optimization can be solved using different nonlinear programming methods. The allocations of water considering different relative weights for the various agencies are presented in Table 7.7. In case 1, domestic demands having the highest priority and industrial demands having less volume have been completely supplied but the utilities of the other agencies are less than 1. Comparing cases 1 and 2 reveals the effect of relative weights on allocated water. Increasing the relative weights of the Department of Water Supply, the Department of Environmental Protection, and the Department of Agriculture has improved the allocation of water to agricultural demands and groundwater table variation, but it does not have any significant effect on the TDS concentration of groundwater.

TABLE 7.7
Results of the Conflict Resolution Method

Variable	Result (Case 1)	Result (Case 2)
Water allocated to agricultural demands (million m^3)	492	533
Water allocated to domestic demands (million m^3)	500	460
Water allocated to industries (million m^3)	80	80
Groundwater table drawdown (m)	2.64	2
Final average TDS concentration of groundwater	1440	1439

7.7 PROBLEMS

7.1 An aquifer (see figure below) is used to supply 3000 m^3/day to industrial demands; the system parameters are presented in following table:

Zone	Storage Coefficient	K_x (m/day)	K_y (m/day)
1	0.02	20	35
2	0.04	45	40

The initial head of the aquifer is 1000 m. The nodes are located in the southern boundary along the river are considered as constant head and other boundaries are impermeable. Develop a simulation model for this system using the Galerkin finite-element method. Calculate the piezometric surface after a 1-month period of operation for the following conditions.

(a) The water withdrawal rates are uniform.
(b) All the water demand is withdrawn from well 3.

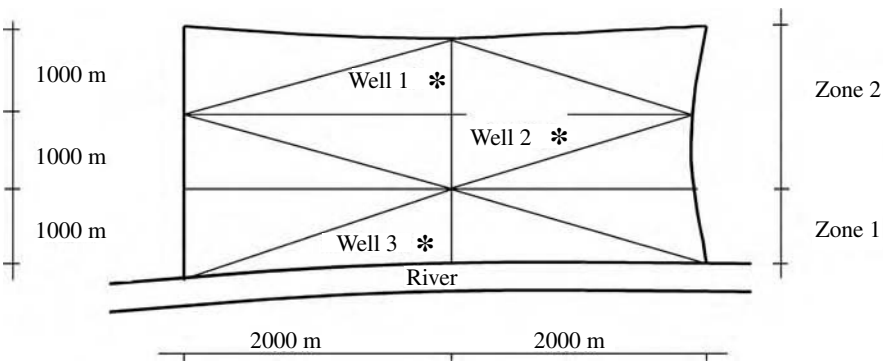


FIGURE 7.1P

7.2 Solve problem 7.1 using implicit and explicit finite-difference methods with a rectangular grid; compare the results and computational costs.

7.3 In the system shown in the following figure, the well is 0.5 m in diameter and fully penetrates a confined aquifer of 40-m thickness and has a hydraulic conductivity of 22 m/day. If the pumping rate is 4 m^3/sec, determine the drawdown at point x and at the well face.

7.4 In the confined aquifer system shown in the following figure, the pumping rates in wells 1 and 2 are 0.02 and 0.03 m^3/sec, respectively. The initial piezometric head is 20 m everywhere. Applying the alternate-direction implicit procedure, determine the head distribution after one day of pumping.

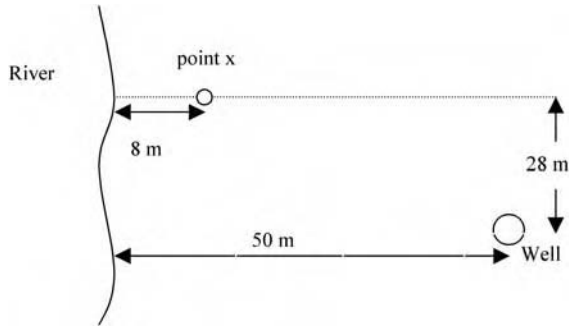


FIGURE 7.3P

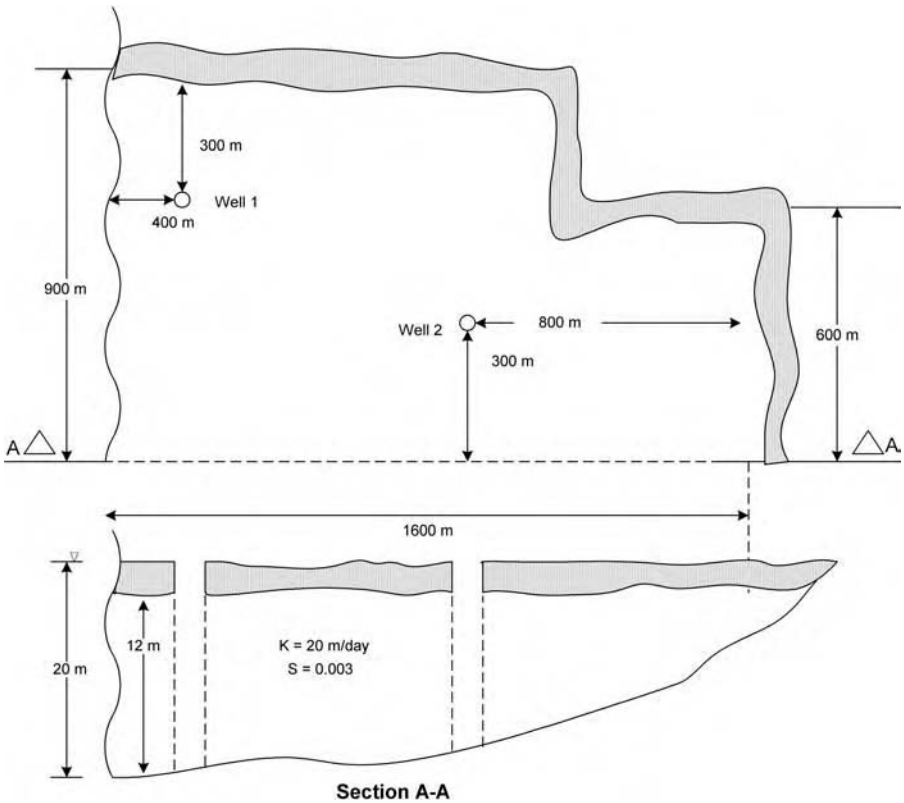


FIGURE 7.4P

7.5 Find the position of the water table, height of the seepage face, and flow rate through a dam with a width of 6 m (see figure below) using a finite-element method. The hydraulic conductivity in x and y directions are 0.001 m/day.

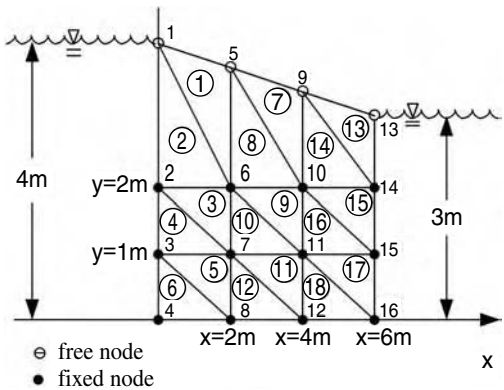


FIGURE 7.5P

7.6 Solve Example 7.9 considering the quality of allocated water. Assume that the average TDS concentration of the groundwater in zones 1 to 4 are 500, 745, 700, and 500 mg/l, respectively. The TDS concentration in Rivers 1 and 2 are provided in the following table:

TDS Concentration in River 1 (mg/l)	TDS Concentration in River 2 (mg/l)
4340	1440
2000	2040
1790	720
370	390
120	450
610	310
2250	1540
2590	1860
3290	1350
1850	2680
3660	1710
2760	1780

The allowable concentration of allocated water is considered to be equal to 1100 mg/l.

7.7 Find the most appropriate water allocation scheme for Example 7.10 using Nash's bargaining solution, considering that 40% of the allocated water returns to the aquifer as return flow and the average TDS concentration of the return flow is 2400 mg/l.

REFERENCES

Cooley, R. L., *A Modular Finite-Element Model (MODFE) for Areal and Axisymmetric Groundwater Flow Problems,* Part 2: *Derivation of Finite-Element Equations and Comparisons with Analytical Solutions*, Book 6, Techniques of Water Resources Investigations series of the U.S. Geological Survey, 1992, chap. A4.

Delleur, J. W., Ed., *The Handbook of Groundwater Engineering*, CRC Press/Springer-Verlag, Boca Raton, FL, 1999.

Esling S. P., Larson, T. A., and Sharpe, D. M., *Graphic Groundwater User's Manual*, Micro-Innovation, Inc., Carbondale, IL, 1993.

Freeze, R. A. and Cherry, J. A., *Groundwater*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

Gupta R. S., *Hydrology and Hydraulic Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

Hantush, M. S. and Jacob, C. E., Non-steady radial flow in an infinite leaky aquifer, *Trans Am. Geophys. Union*, 36(1), 95–100, 1955.

Harr, M. E., *Groundwater and Seepage*, McGraw-Hill, New York, 1962.

Istok, J., *Groundwater Modeling by the Finite Element Method*, AGU Water Resources Monograph 13, Washington DC, 1989.

Karamouz, M., Kerachian, R., Zahraie, B., and Zahraie, B., Conjunctive use of surface and groundwater resources, in *Proc. ASCE Environmental and Water Resources Institute Conf.*, Roanoke, VA, May 19–22, 2002.

Kashef, A. A. I., *Groundwater Engineering*, McGraw-Hill, New York, 1987.

Lee, C. H., The determination of safe yield of underground reservoirs of the closed basin type, *Trans. Am. Soc. Civ. Eng*., 78, 148–151, 1915.

McDonald, M. G. and Harbaugh, A. W., *A Modular Tree-Dimensional Finite-Difference Groundwater Flow Model*, Book 6, Techniques of Water Resources Investigations series of the U.S. Geological Survey, 1988, chap. Al.

ReVelle, C. and McGarity, A. E., Eds., *Design and Operation of Civil and Environmental Engineering Systems*, John Wiley & Sons, New York, 1997.

Todd, D. K., *Groundwater Hydrology*, John Wiley & Sons, New York, 1980.

Torak, L. J., *A Modular Finite-Element Model (MODFE) for Areal and Axisymmetric Groundwater Flow Problems*, Part 1: *Model Description and User's Manual*, Book 6, Techniques of Water Resources Investigations series of the U.S. Geological Survey, 1993, chap. A3.

Willis R. and Yeh, W. W.-G., *Groundwater Systems Planning and Management*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

Yeh, W. W.-G., System analysis in groundwater planning and management, *J. Water Resources Plan. Manage.*, 118(3), 224–237, 1992.

Zienkiewicz, O. C., *The Finite-Element Method in Engineering Science*, McGraw-Hill, New York, 1971.

8 River–Reservoir Systems Modeling

8.1 INTRODUCTION

This chapter discusses elements of river–reservoir systems modeling, including data and information processing, objectives, constraints identification, and details of the modeling process. Also presented are the objectives of river–reservoir systems planning and management modeling, including water supply, flood control, power generation, and supplying instream flow requirements, as well as formulation of simulation and optimization models for different objectives.

8.2 DATA COLLECTION AND PROCESSING

8.2.1 Inflow

Inflow to reservoirs is usually measured at hydroclimatic stations located on tributaries entering the river and reservoirs and other control points. In some of the reservoirs, the inflow is measured by the water balance in the reservoir using the actual release figures on an hourly or daily basis. Streamflow has to be estimated when there is no direct observation is lacking. Inflow to the reservoir is one of the most important sources of uncertainty in development of operating policies for a system. In a systems approach to reservoir operation and modeling, it is preferable to have a long record of streamflows that includes worst-case scenarios of droughts and floods experienced during the historical record.

Deterministic models for reservoir operation and modeling for expansion of the system do not implicitly consider inflow uncertainties; however, they have been widely used for stochastic modeling of complex and large-scale water resources systems. In these models, the uncertainty is explicitly incorporated using generated synthetic time series of uncertain inflows. Statistical models such as ARMA (autoregressive moving average) have been widely used by investigators to generate long records of data. In stochastic modeling of reservoirs, the statistical behavior and distribution of historical records is used in order to model the uncertainty of inflows.

Seasonal stochastic processes are frequently applied to hydrologic time series in order to model the data affected by stochastic periodic components. This approach allows one to fit many records that exhibit seasonal patterns caused by the periodicity in the data (Montanari et al., 2000).

Defining hydrologic seasons is an important step when estimating forecast parameters and defining heuristics for seasonal and monthly forecasts. Statistical modeling of hydrologic time series can be improved significantly by defining seasons

in a region (Karamouz, 1999a). Seasons could be defined based on a number of factors, such as:

- Estimation of the snow budget over the watershed and analysis of the correlation between the average snow water equivalent for different months of snow accumulation or snowmelt seasons and seasonal stream flows
- Similarity in basic statistical characteristics of streamflow data, such as mean and coefficient of variation for different months of a season
- General hydrologic characteristics of the study area, such as variations in precipitation, snow budget, and groundwater contribution to surface water recharge
- Other factors, such as period of snowmelt and streamflow associated with specific climatic regimes such as monsoon and tropical storms

The main objective of defining seasons is to capture the coupled effect of climatic and hydrologic factors. It also allows modelers to reduce the number of estimated parameters (compared with monthly time steps). Models with fewer parameters to estimate tend to perform better and have a smaller tendency for error accumulation (parsimony of parameters) (Salas et al., 1982).

Some of the stochastic reservoir operation models and their structures are explained in the following text. In these models, inflow forecast time series are utilized to incorporate uncertainty in inflows and the seasonal forecast model itself. In addition to modeling uncertainties, inflow forecasts are necessary for applying optimal operation policies to real-time operation of reservoirs.

A number of forecast models have been used for reservoir inflow forecasting for real-time operation. These methods follow approaches ranging from pure statistical methods to those based primarily on conceptual modeling of the hydrologic cycle. Development of statistical models started with some very simple regression models and significantly improved to incorporate seasonal and nonseasonal correlations. Details of some of the most widely used statistical models are presented in Chapter 5.

Figure 8.1 shows the data required and steps that should be taken for reservoir inflow modeling and forecasting. As can be seen in this figure, the algorithm utilizes a wide range of hydrologic and climatic data. Using such data will significantly improve the accuracy of a forecast model (Karamouz et al., 2001). Different forecast modification schemes are also considered in the algorithm and are used to reduce the errors of statistical models.

8.2.2 Water Demands

Different water demands and their priorities are the main constraints in reservoir operation optimization problems. Water-use circulars prepared by the U.S. Geological Survey (USGS) identify the following water uses in the overall water accounting system (Solley et al., 1993):

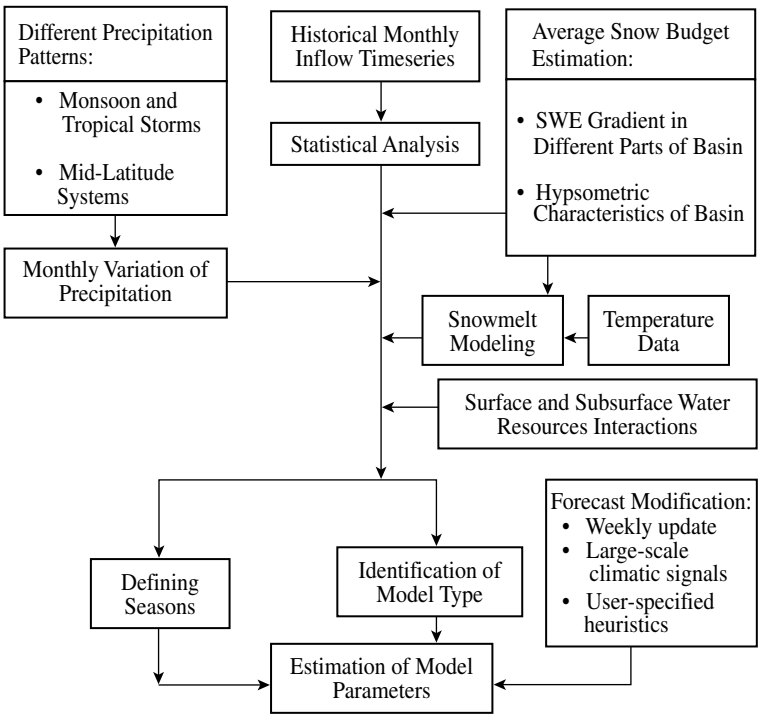


FIGURE 8.1 An algorithm for reservoir inflow modeling.

- Water withdrawal for offstream purposes
- Water deliveries at point of use or quantities released after use
- Consumptive use
- Conveyance loss
- Reclaimed wastewater
- Return flow
- Instream flow

Water demands in river–reservoir systems resulting from offstream uses include:

- Domestic or municipal
- Industrial
- Agricultural

These uses require withdrawal of water from the surface or groundwater resources. Part of the water withdrawn may return to the system, perhaps in another location with different quality and over a different period of time. Water can also be allocated to instream uses that alter the distribution of flow in time and space (Loucks et al., 1981). Such uses include:

- Reservoir storage, possibly for recreational uses or navigation
- Flow augmentation for water quality control or for navigation
- Hydroelectric power production

There is a lag time between water demands and inflows. The peak of water demand may occur in months during which inflows are at a minimum such as in the middle of summer. The main purpose of water resources modeling is to derive policies for water allocation in a way that demands can be met with reasonable certainty. Thus, to manage a system, it is important to be able to describe the current and projected water use fluctuations and variations.

Water demands change from year to year and month to month. Many physical, economical, social, and political reasons for these variations can be identified. In recent years, significant climatic changes have been observed in many parts of the world, including more intense floods, greater precipitation, and even unusual droughts in some regions. These changes have significantly affected the water demands in many parts of the world.

8.2.3 Physical Characteristics

8.2.3.1 Types of Dams

Dams are usually classified in terms of materials and forms. Common types are homogenous or zoned earthfills, rockfills with an earth core or concrete face, and concrete dams. Concrete dams are classified as gravity, arch, and buttress resistance (Figure 8.2). Topography and geomorphology are primary factors in weighting the comparative merits of dam types.

Embankment dams are constructed of earth and/or rock with a provision for controlling seepage by means of an impermeable core, concrete face, grouting curtain, or upstream blanket. Embankment dams have relatively poor resistance to overflow; therefore, the spillway capacity must be determined conservatively.

Arch dams can carry large loads, but their integrity depends inherently on the strength of abutments. A gravity dam is an essentially solid concrete structure that resists imposed forces principally by its own weight. Although they are usually straight in plan, these dams are sometimes curved or angled to accommodate site topography.

A buttress dam is a gravity structure that, in addition to its own weight, utilizes the weight of water over the upstream face to provide stability. The simplest type of these dams is the slab type, which consists of sloping flat slabs supported at intervals by buttresses.

Storage capacity is the most important physical characteristic of reservoirs. The storage can be determined for each level of water from topographic map of the site. An area–volume–elevation curve can be constructed by implementing the area enclosed with each topographic contour in the reservoir site and summation of the increments of storage below each level. This curve can be used in selection of total capacity for reservoir and reservoir operation optimization. Figure 8.3 shows an area–volume–elevation curve.

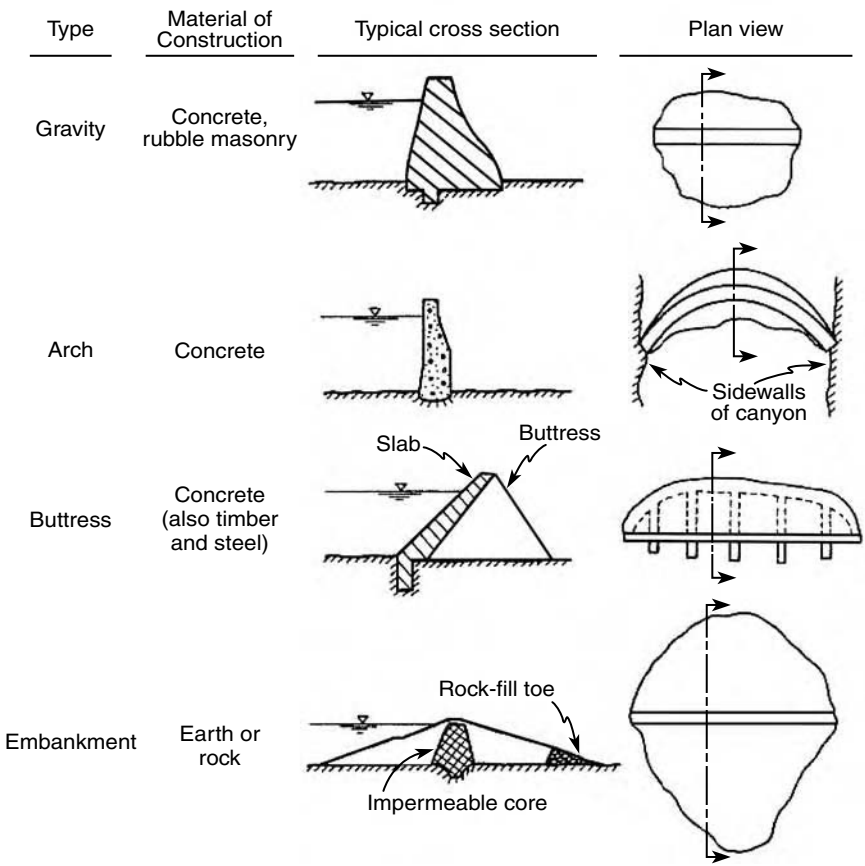


FIGURE 8.2 Basic types of dams. (From Linsley, R. K. et al., *Water Resources Engineering*, McGraw-Hill, New York, 1992. With permission.)

8.2.3.2 Ancillary Facilities

A *spillway* is a safety valve for the dam and should have enough capacity to discharge major floods without damaging the dam and ancillary structures and also keep the reservoir level below the maximum water level. A spillway that has gates to adjust the rate of outflow is referred to as a *controlled* spillway. An *uncontrolled* spillway will discharge water based on the water level in the reservoir.

The majority of the active storage of reservoirs is below the spillway crest, and *outlets* must be built in order to withdraw water from various parts of the reservoir storage. Release from these outlets may be discharged into a channel below the dam or may be transported in pipes or canals to some distant point. The basic structural components of outlets are *sluiceways* and *intakes*. A sluiceway is a pipe or tunnel that passes through a dam or hillside at one end of the dam and discharges into the stream. An intake structure is required at the entrance to a conduit through which

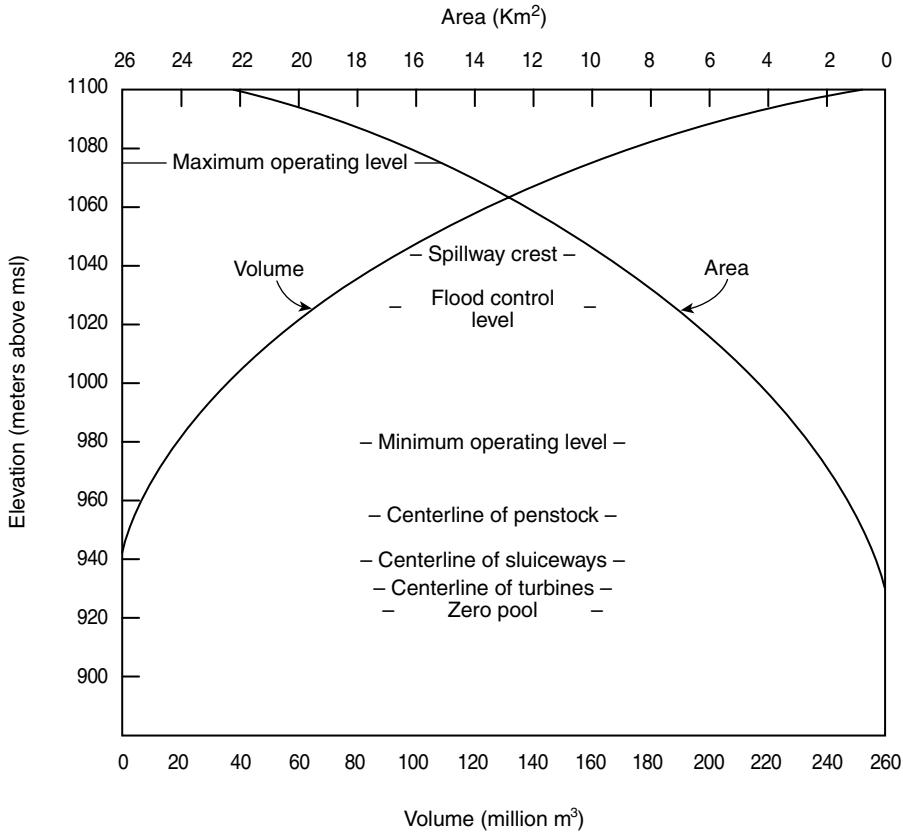


FIGURE 8.3 An example of area–volume–elevation curve of dams.

water is withdrawn from the reservoir. Large dams may have sluiceways and/or intakes in different levels (Linsley et al., 1992).

8.2.3.3 Hydropower Facilities

Listed here are the basic elements of a hydropower plant (see Figure 8.4) (Mays, 1996):

- Reservoir for creating the necessary head that will provide energy needed for deriving turbines
- Intake structure that directs water from the reservoir into the penstock or conduit; gates or valves are used to control water discharge to the power plant, and racks or screens prevent trash or debris from entering the turbines
- Conduit that conveys water from the intake structure to the powerhouse
- Power plant, including turbines, generators, and control and auxiliary equipment

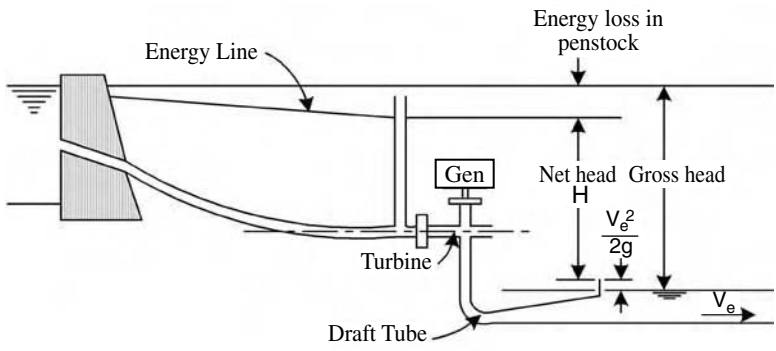


FIGURE 8.4 Element of a typical hydropower plant.

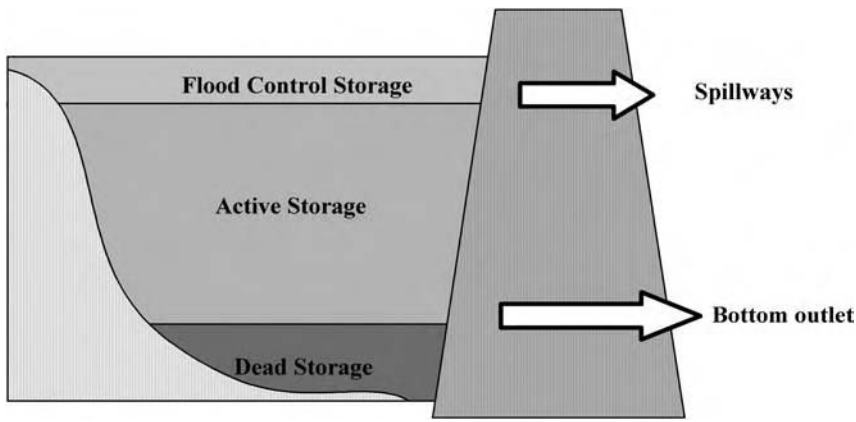


FIGURE 8.5 Reservoir storage zoning.

8.3 OBJECTIVES

Large dams are usually multipurpose structures. Besides providing water for domestic, agriculture, and industrial uses (the main objectives of reservoir planning and operation), hydropower electric production is another objective of development of many river–reservoir systems. High efficiency, lower costs, and the specific capabilities of hydropower plants for controlling the frequency of power networks have made hydropower plants a necessary component of power systems. Flood control and damage reduction is another objective for dam construction. A reservoir reduces the peak flow of a flood hydrograph to an amount lower than the river carrying capacity. In the following sections, these objectives are briefly explained and the structure of reservoir optimization models for satisfying demands is demonstrated. Navigation requirements and environmental water rights are also presented. Other purposes, such as quality control in the reservoir and downstream river, are

thoroughly explained in Chapter 9. Modeling power generation systems is also explained in detail in Chapter 10.

8.3.1 Water Supply: Reservoir Storage Zones

As shown in Figure 8.5, the total capacity of the reservoir can be divided into three major parts:

- Active storage
- Dead storage
- Flood control storage

Active storage is the volume between minimum and normal water levels. Normal water level is the maximum elevation of a reservoir during normal operation. For many reservoirs, the normal water level is the elevation of the spillway crest, and the minimum water level is the lowest elevation of the reservoir during normal operation and may be fixed by elevation of the lowest outlet or minimum head required for hydropower generation. Active storage is required for conservation purposes, including water supplies, navigation, etc.

Flood control storage is reserved for storing excessive flood volumes to prevent overtopping of the dam and to reduce the potential flood damage. The flood control storage zone is between normal water level and maximum water level in the reservoir. During floods, discharge over the spillway may cause the water level to rise above normal water level. This excess storage is *surcharge storage* and is normally uncontrolled.

8.3.2 Power Generation

Energy production is a major objective of water resources development in many river–reservoir systems. In hydropower plants, energy is produced when water of adequate head runs electric turbines and generates electricity without actually consuming water. As can be seen in Figure 8.4, the dam provides the required head. The efficiency of power generation in the power plant shown in Figure 8.4 depends upon:

- Plant efficiency
- Volumetric water flow through the turbine
- Net hydraulic head of water above the turbine

The energy production of a hydropower plant could be estimated using the following equation (Mays, 1996):

$$EO = \int e_f \, \gamma \, Q(t) \, H(t) \, dt \tag{8.1}$$

where EO is the electrical output of power plant, e_f is the overall efficiency of power generation, and γ is the specific weight of water. As can be seen in this equation,

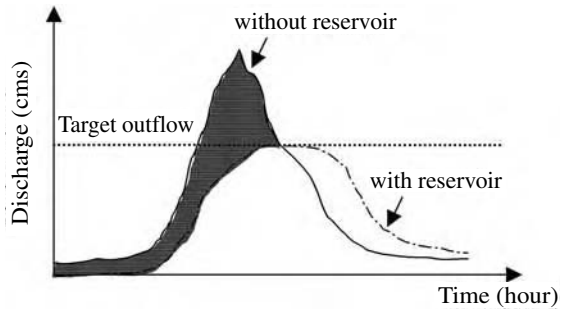


FIGURE 8.6 Impact of storage on flood peak discharge reduction.

increasing head, $H(t)$, and discharge, $Q(t)$, through the turbines increases the energy production. Efficiency is also a function of net hydraulic head above the turbines.

8.3.3 Flood Control

To reduce flood damage, part of the active storage of the reservoirs is reserved for flood control. As it can be seen in Figure 8.6, the volume of water represented by the shaded area is stored and then released gradually at a rate that does not exceed the carrying capacity of the river. The total volume of the inflow and outflow hydrographs is almost the same (less evaporation) but the time distribution is changed by the reservoir.

8.3.4 Navigation Requirements

Rivers and lakes have long been used for public and commercial transportation. Recreational boating is also another concern in the operation of some river–reservoir systems. The main constraint for potentially navigable parts of the river downstream of a reservoir is maintaining a minimum acceptable depth at various sites s ($EL_s^{\min}$). Assuming the known river flow–stage relationship (FS_t^s), the constraint that should be considered in the reservoir operation optimization model is as follows (Loucks et al., 1981):

$$FS_t^s\left(EL_t^s\right) \geq EL_s^{\min} \tag{8.2}$$

where (EL_t^s) is the depth in time step t.

A more detailed analysis of navigation benefits and operation policies may be necessary, especially when locks are required. If capital investment in locks is necessary, their fixed costs as a function of the design depth of the channel, any expected operating costs at each lock site s, and benefits of the transportation system should be included in the objective function of the planning models.

8.3.5 Environmental Water Rights

Providing adequate water to protect the aquatic, biological, and aesthetic values of a stream and to preserve existing fisheries is an important constraint in river–reservoir

systems planning and operation. Because of the variety of habitats and streamflow quality and quantity conditions, estimation of minimum instream flow requires comprehensive studies. The simplest constraint that can be considered for satisfying necessary environmental conditions is to supply the minimum monthly flow required in each reach of the river ($Q_{\min}^{NM}$) after supplying demands:

$$R_t - \sum_{j=1}^{m} RA_{j,t} \geq Q_{\min}^{NM} \qquad (t = 1, \cdots, n)\ (NM = 1, \ldots, 12) \tag{8.3}$$

where R_t is release from reservoir in month t and $RA_{j,t}$ is water allocated to demand j in month t. NM is the number of months and m is the total number of water demands.

8.4 CAPACITY EXPANSION

Capacity expansion is mainly focused on increasing water supply by structural methods. Construction of a new dam, increasing the height of an existing reservoir, and construction of tunnels or channels for interbasin water transfers are some common examples of capacity expansion; however, they are somewhat limited by site-related, geophysical, geographic, economic, and institutional factors.

8.4.1 Interbasin Water Transfer

Historically, many attempts have been made to divert water from areas of water surplus to areas of water deficit. *Interbasin water transfer* refers to water withdrawal and removal by means of ditches, canals, tunnels, or pipelines from a source containing excess water to an adjacent area experiencing a shortage. The shortage capacity in the importing basin is the main constraint in determining the design capacity of water transfer structures, but economic and geophysical constraints are also important limitations. The following model is the general form of an optimization model that can be used for optimizing the design capacity based on economic analysis and the projected life of the project:

$$\text{Minimize} \qquad z = \sum_{t=1}^{n} \left(C_t - B_t \right) \tag{8.4}$$

$$\text{Subject to:} \qquad Q_d \leq Q_M \tag{8.5}$$

$$C_t = f_t(Q_d) + C'_{i,t} + C_{e,t} \quad \forall t \tag{8.6}$$

$$B_t = B_{i,t} - B'_{e,t} \quad \forall t \tag{8.7}$$

where:

n is the planning horizon (time period).
Q_d is the design capacity of water transfer system.
Q_M is the maximum capacity of the water transfer system (based on geophysical characteristics of the study area and supply limitations).
C_t is the present value of the total cost of the system after implementation of water transfer system in time period t.
$f_t(Q_d)$ is the present value of the total cost of the water transfer system including capital investment, operation, and maintenance costs in time period t.
$C'_{i,t}$ is the present value of the cost of importing water from basin i after implementation of the water transfer system in time period t.
$C_{e,t}$ is the present value of the costs of exporting water to basin e due to water transfer in time period t.
B_t is the present value of benefits of the system after implementations of the water transfer system in time period t.
$B_{i,t}$ is the present value of the net benefits of importing basin i in time period t.
$B'_{e,t}$ is the decrease in the present value of the net benefits of exporting basin e in time period t.

In order to estimate the present value of an investment in year t of the planning time horizon, the following equation can be used:

$$P = \frac{F}{(1+i')^t} \tag{8.8}$$

where P and F are the present and future value, respectively, of the investment in year t; i' is the interest rate (for more details, see Chapter 4).

Example 8.1

Based on a population growth analysis, the water demand of a city is estimated to change over time as shown in Figure 8.7. Four interbasin water transfer projects are studied for supplying future demands of this city over a 20-year planning time horizon. The capacity and initial investment of these projects are shown in Table 8.1. The interest rate is considered to be 6%. Find the optimal sequence of implementing projects.

Solution: Table 8.2 shows the present value of the initial investment in the projects which has been estimated using Eq. (8.8) for the years in which water demand has increased. In order to find the best timing for construction of interbasin water transfer projects, minimization of the present value of the costs is considered as the objective function of a linear programming model as follows:

Minimize $C = 1.78X_{11} + 1.12X_{12} + 0.88X_{13} + 0.62X_{14}$
$+ 2.22X_{21} + 1.40X_{22} + 1.11X_{23} + 0.78X_{24}$
$+ 1.60X_{31} + 1.01X_{32} + 0.80X_{33} + 0.56X_{34}$
$+ 3.56X_{41} + 2.23X_{42} + 1.77X_{43} + 1.25X_{44}$

Subject to: $15X_{11} + 20X_{21} + 10X_{31} + 30X_{41} \geq 20$
$15X_{12} + 20X_{22} + 10X_{32} + 30X_{42} \geq 30$
$15X_{13} + 20X_{23} + 10X_{33} + 30X_{43} \geq 20$
$15X_{14} + 20X_{24} + 10X_{34} + 30X_{44} \geq 10$
$X_{11} + X_{12} + X_{13} + X_{14} \leq 1$
$X_{21} + X_{22} + X_{23} + X_{24} \leq 1$
$X_{31} + X_{32} + X_{33} + X_{34} \leq 1$
$X_{41} + X_{42} + X_{43} + X_{44} \leq 1$

$$X_{ij} = \begin{cases} 1 & \text{if the project i constructed in the j}^{\text{th}} \text{ time period} \\ 0 & \text{otherwise} \end{cases}$$

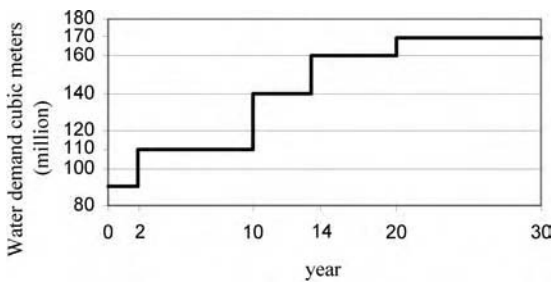


FIGURE 8.7 Variation in water demand of a city over a 30-year period.

TABLE 8.1
Initial Investment and Capacity of Interbasin Water Transfer

Interbasin Water Transfer Project	Initial Investment (10^6)	Capacity (million m^3)
A	2	15
B	2.5	20
C	1.8	10
D	4	30

TABLE 8.2
Present Value of Initial Investment for Interbasin Water Transfer Projects

Project	t = 2 years	t = 10 years	t = 14 years	t = 20 years
A	1.78	1.12	0.88	0.62
B	2.22	1.40	1.11	0.78
C	1.60	1.01	0.80	0.56
D	3.56	2.23	1.77	1.25

In this problem, the constraints are defined in order to supply demands in different years and to ensure that each project can only be constructed once. Optimal timing for implementation of the projects is: build projects C at year 2, D at year 10, B at year 14, and A at year 20. The present value of the costs of construction of the projects is estimated to be 5.56 million dollars.

8.4.2 Optimal Sizing of a Single Reservoir System

If the water demand of the river–reservoir system remains constant in different months, then a basic planning model searches for the minimum reservoir capacity required for supplying water with a steady release, which could be formulated as follows:

$$\text{Minimize} \quad z = Cap \tag{8.9}$$

$$\text{Subject to:} \quad S_{t+1} = S_t + I_t - R - E_t - L_t \qquad (t = 1,\ldots,n) \tag{8.10}$$

$$S_{\min} \le S_t \le Cap \qquad (t = 1,\ldots,n) \tag{8.11}$$

$$D \le R \le R_{\max,t} \qquad (t = 1,\ldots,n) \tag{8.12}$$

$$S_t, E_t, L_t, R \ge 0 \qquad (t = 1,\ldots,n) \tag{8.13}$$

where Cap is the total capacity of the reservoir, and S_t and I_t are storage at the beginning of month t and inflow to the reservoir during month t, respectively. The main constraint of the model is the continuity equation. As shown in Eq. (8.10), storage at the end of each month is equivalent to the storage at the beginning of the month plus inflow minus release, evaporated water (E_t), and leakage from the reservoir (L_t). The volume of water evaporated is a function of the evaporation rate and the surface area of the reservoir. Equation (8.11) shows the upper and lower bound on the volume ($S_{\min}$). Equation (8.12) limits any release from the reservoir to less than the maximum allowable release ($R_{\max,t}$) (such as the carrying capacity of the river) downstream of the reservoir, depending on the capacity of the hydraulic structures (e.g., penstock or spillway). The minimum release is assumed to be a constant demand, D, but a fraction of demand could be used as minimum release if water allocation could be relaxed.

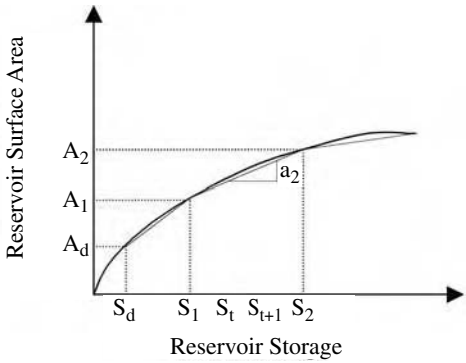


FIGURE 8.8 Reservoir storage–area relationship and piecewise approximation.

As shown in Figure 8.8, the evaporation losses can be estimated using piecewise approximation of the reservoir area–volume curve. The evaporation losses could be calculated using the information from this figure:

$$E_t = a_2 \cdot e_t\left(\frac{S_t + S_{t+1}}{2}\right) + A_1 \cdot e_t \tag{8.14}$$

where e_t is the rate of evaporation (m) in period t and a_2 is the slope of segment 2. Equations (8.9) to (8.14) are linear; therefore, this problem can be solved with linear programming. Based on the above formulation, the optimization algorithm will search for the minimum capacity required to supply the total demand all the time. But, in many real cases, such a scenario is not economically feasible. In other words, based on economic analysis of the costs and benefits of the system, the planner may forego meeting the demands and choose a smaller capacity that could cause shortages in critical periods of operation. For this purpose, the operating policy is sought in some conjunction with the search for optimal reservoir capacity (Mays and Tung, 1992). The objective function of simultaneous optimization of operation and sizing of the reservoir can be formulated as follows with the same constraints as for Eqs. (8.10) to (8.13):

$$\text{Minimize} \quad z = \sum_{t=1}^{n}\left(C_t - B_t\right) \qquad (t = 1,\ldots,n) \tag{8.15}$$

As seen in Eq. (8.15), the objective function is considered to be the net revenue of the system. The costs of operation can be classified as follows:

- Initial investment
- Operation and maintenance costs
- Costs due to shortages in supplying water demand
- Costs due to damages incurred during emergency situations, such as floods and droughts

Linear programming techniques cannot be used for solving this problem because the costs and benefits of operating the dams are usually nonlinear functions of parameters such as reservoir release compared with water demands. Nonlinear optimization and dynamic programming techniques are frequently used to solve this problem. Details of these optimization techniques were explained in Chapter 2, and application of these methods for optimization of reservoir operation and design is explained later in this chapter.

In real-life situations, release from reservoir varies from month to month (R_t) and reservoirs supply water to varying demands. In these cases, formulation of the reservoir sizing problem would be as follows:

$$\text{Minimize} \quad z = \sum_{t=1}^{n} \left(C_t - B_t\right) \tag{8.16}$$

$$\text{Subject to:} \quad S_{t+1} = S_t + I_t - R_t - E_t - L_t \qquad (t = 1, \ldots, n) \tag{8.17}$$

$$S_{\min} \le S_t \le Cap \qquad (t = 1, \ldots, n) \tag{8.18}$$

$$0 \le R_t \le R_{\max,t} \qquad (t = 1, \ldots, n) \tag{8.19}$$

$$S_t, E_t, L_t, R_t \ge 0 \qquad (t = 1, \ldots, n) \tag{8.20}$$

In this case, the release from reservoir would be allocated to various demands:

$$R_t = \sum_{j=1}^{m} RA_{j,t} + Q_t \qquad (t = 1, \ldots, n) \tag{8.21}$$

where $RA_{j,t}$ represents water allocated to demand j in month t and Q_t represents water remaining in the downstream river. The allocation policies depend on contractual, legal, and institutional requirements for various purposes.

8.4.3 Reliability-Based Design

Usual design practice is based on recommended factors of safety, which should allow the system to be subject to only a limited number of failures over its projected lifetime. This approach is usually based on a maximum probability of failure within a given lifetime (n) of the system.

Based on this approach and taking into account economic aspects such as project costs, benefits, or benefit–cost ratio, the optimal size of a reservoir can be estimated. In reservoir sizing problems, it is assumed that the probability distribution of inflows is known. As the design return period increases, the initial cost of the dam and accompanying facilities increases while the expected damages associated with

unsatisfactory operation of the system decreases. For each return period, the total cost can be estimated by adding the expected annual damage costs and the capital costs of constructing a reservoir that can control the floods with that return period. Then, the design return period can be found by estimating the minimum total cost. Suppose X is the variable for maximum annual flood of the reservoir. If the design return period of a reservoir is considered to be T, then a flood with magnitude x will incur no cost if $x \leq x(T)$. The expected annual cost can be estimated as (Kottegoda and Rosso, 1997):

$$d_T = \int_{x(T)}^{+\infty} d(x)\, f_x(x)\, dx \tag{8.22}$$

where $f_x(x)$ is the probability density function of floods and $d(x)$ is the cost function and $x(T)$ is the magnitude of a flood with return period T.

Example 8.2

The design return period T of a dam that is facing floods can be estimated using a reliability-based approach. As can be seen in Table 8.3, the initial cost or capital investment required for a dam and the facilities increases as the design return period increases. It is assumed that the probability distribution of the floods and cost of damages can be estimated. As shown in the table, the damage cost increases as the flood return period increases. Find the optimal design return period using the reliability-based approach. A finite-difference approximation of Eq. (8.22) is used as follows:

$$\text{Incremental expected damage} = \frac{d\left(T_{i-1}\right) + d\left(T_i\right)}{2}\left(\frac{1}{T_{i-1}} - \frac{1}{T_i}\right)$$

where $d(T_i)$ is the estimated damage if the flood with return period T_i occurs. The risk of damage for each return period is then computed using partial summation of the relevant incremental values. The total costs are computed by adding the corresponding capital investment. As shown in Table 8.3, the minimum cost is estimated for a 500-year return period. The risk of failure for the design based on this return period is 0.002.

8.4.5 Sedimentation and Dredging

The sedimentation rate for a reservoir is an important factor in determining the active storage and useful life of the reservoir. The projected rate of reduction in useful storage of an existing reservoir is a significant constraint in capacity expansion problems. The main issue is to decide whether to construct a new dam, modify the land use and vegetation upstream of the dam to reduce sediment transport to the reservoir, or to remove the accumulated sediments from the reservoir bed by mechanical means such as dredging.

TABLE 8.3
Calculation of Optimum Design Capacity of a Reservoir and Facilities

Return Period (years)	Annual Probability of Exceedance	Capital Cost ($10^6/year)	Damage ($10^6)	Incremental Expected Damage ($10^6/year)	Expected Damage Cost ($10^6/year)	Total Cost ($10^6/year)
100	0.01	20.1	0	—	53.8	73.9
200	0.005	21.2	500	1.25	52.55	73.75
500	0.002	32.1	10,000	15.75	36.8	68.9
1000	0.001	45.4	15,000	12.5	24.3	69.7
10,000	0.0001	70.5	39,000	24.3	0	70.5
Sum				53.8		

8.5 RESERVOIR OPERATION OPTIMIZATION

8.5.1 Water Supply with Monthly Varying Release

The basic model for reservoir operation when the releases from the reservoir are varying from month to month is as follows:

$$\text{Minimize} \quad z = \sum_{t=1}^{n} loss_t\left(R_t, D_t, S_t\right) \tag{8.23}$$

$$\text{Subject to:} \quad S_{t+1} = S_t + I_t - R_t - E_t - L_t \qquad \left(t = 1, \ldots, n\right) \tag{8.24}$$

$$S_{\min} \le S_t \le Cap \qquad \left(t = 1, \ldots, n\right) \tag{8.25}$$

$$0 \le R_t \le R_{\max,t} \qquad \left(t = 1, \ldots, n\right) \tag{8.26}$$

$$S_t, E_t, L_t, R_t \ge 0 \qquad \left(t = 1, \ldots, n\right) \tag{8.27}$$

As it can be seen in Eq. (8.23), the objective function of the model is to minimize total losses in the historical period of operation. Loss functions have been proposed to be a function of shortages in supplying demands and the difference between water storage in the reservoir and target storage in each month.

Based on the above formulation, total losses are estimated by comparing total releases with total water demands of the system. The optimal releases for allocating water to different demands are determined by utilizing a simulation model; in this approach, the details of supplying shortages are not considered within the optimization routine. In a more elaborate algorithm for multipurpose reservoir operation,

the allocation policy is applied in the optimization loop. In other words, each time the optimization algorithm searches for the optimal solution, the losses are estimated based on the details of water allocation to the different demands. The objective function of this model would be as follows:

$$\text{Minimize} \quad z = \sum_{t=1}^{n} \sum_{j=1}^{m} loss_t\left(R_{j,t}, D_{j,t}, S_t\right) \tag{8.28}$$

where:

$R_{j,t}$ is water allocated to demand j in month t.
$D_{j,t}$ is the volume of water demand j in month t.
S_t is reservoir storage in month t.
m is the number of water demands.

The constraints are the same as before.

Example 8.3

A reservoir is constructed on a river for supplying the power load of a city located near the river and water demands of downstream agricultural lands. The water that is released for power generation can also be used for supplying other demands. The required instream flow is estimated to be 1 million m^3 each month. The total capacity of the reservoir is 10 million m^3. Maximum monthly release from the reservoir is limited to 7 million m^3. Table 8.4 shows the monthly inflows to the reservoir and benefits of power generation and water supply. The reservoir storage on January 1 is considered to be 5 million m^3. Find the optimal releases using linear programming.

Solution: The following formulation can be used for finding the optimal monthly volume of water that should be released from the reservoir:

$$\text{Maximize} \quad B = 2.6\,R_1 + 2.9\,R_2 + 3.6\,R_3 + 3.9\,R_4 + 4.2\,R_5 + 4.2\,R_6 + 4.5\,R_7 + 4.1\,R_8 + 3.6\,R_9 + 3.1\,R_{10} + 2.7\,R_{11} + 2.5\,R_{12}$$

$$\text{Subject to:} \quad S_1 = 5$$

$$S_i + I_i - R_i = S_{i+1} \qquad (i = 2, \cdots, 12)$$

$$1 \le R_i \le 7 \qquad (i = 1, \cdots, 12)$$

$$S_i \le 10 \qquad (i = 1, \cdots, 13)$$

By solving this problem using a linear programming package, the total maximum benefit is estimated as \$128,000 for the two sets of optimal monthly releases shown in Table 8.5.

TABLE 8.4
Monthly Inflows to the Reservoir and Benefits of Power Generation and Water Supply in Example 8.3

Month	Inflow (million m^3)	Benefits of Power Generation ($\$10^3$/million m^3)	Benefits of Water Supply ($\$10^3$/million m^3)
January	2	1.6	1.0
February	2	1.7	1.2
March	3	1.8	1.8
April	4	1.9	2.0
May	3	2.0	2.2
June	2	2.0	2.2
July	2	2.0	2.5
August	1	1.9	2.2
September	2	1.8	1.8
October	3	1.7	1.4
November	3	1.6	1.1
December	2	1.5	1.0

TABLE 8.5
Optimal Releases from the Reservoir in Example 8.3

Jan.	*Feb.*	*Mar.*	*Apr.*	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1	1	1	3	7	3	7	1	2	3	3	2
1	1	1	3	3	7	7	1	2	3	3	2

8.5.2 Flood Control

The flood storage capacity of a reservoir is one of the important components of many flood control systems. For this purpose, part of the active storage of a reservoir is kept empty to store potential floods and gradually release excess water at rates not exceeding the capacity of the downstream river. Methods of assessing the volume required for flood storage are based on minimizing expected damages. To incorporate the flood control objective, the following constraint should be considered in reservoir sizing or operation models:

$$S_t \le Cap - V_t \qquad (t = 1, \ldots, n) \tag{8.29}$$

where V_t is the flood control storage in month t. Flood routing procedures, along with knowledge of flood control operating policies and channel storage characteristics, can be used to predict the impact of flood peaks on downstream of the reservoir. Various discrete flood storage capacities in an upstream reservoir could be utilized to provide a buffer for extreme flood situations. The important step for determining

flood control storage for different months is to define the hydrographs that represent the response of the watershed to floods with varying return periods. The following steps may be taken (Karamouz and Zahraie, 1998):

- Identification of observed flood hydrographs of significant magnitude
- Selection of representative hydrographs showing flood characteristics of the watershed
- Estimation of flood peaks of different return periods using comprehensive statistical analysis of observed values and fitting the best statistical distribution
- Extrapolation of flood peaks on the representative flood characteristics

The next step would be to determine the ability of a reservoir to withstand floods of various magnitudes. Physical characteristics of the reservoir, discharge outlets, and the carrying capacity of the river downstream of the reservoir should be taken into account. Figure 8.9 shows the steps that should be taken to determine flood storage necessary for floods of various return periods. As seen, the final result of this algorithm is that the reservoir will be graduated to flood control storage required for floods with varying probabilities of occurrence.

Example 8.4

Consider a dam constructed near a city as shown in Figure 8.10. Reservoir capacity is considered to be 70 million m^3. Control point A, considered to be a critical point within the city, is located 10 km downstream of the reservoir and has a relatively lower carrying capacity than the rest of the reaches of the river. The river carrying capacity at this control point is estimated to be 200 m^3/sec. A local stream discharges to the river upstream of the city. Figure 8.11 shows the inflow to the reservoir and the local flow hydrographs estimated for a 100-year flood. As can be seen in this figure, the peak flow of the inflow to the reservoir is about 280 m^3/sec, which is higher than the river carrying capacity. For finding the flood control volume, variation of the peak reservoir discharge with respect to the reservoir storage level at the beginning of the flood should be investigated.

The HEC-5 package (Hydrologic Engineering Center, 1982) is used to estimate the required flood control storage for reducing a 100-year flood peak to the carrying capacity of the river at the control point. The Maskingum method is used for routing the reservoir outflow and local flows in the river between the dam and control point A. It is assumed that the storage of the reservoir is 67 million m^3 at the beginning of the flood. Reservoir outflow and the control point discharge hydrographs for this scenario are shown in Figure 8.12. The discharge at control point A is higher than the carrying capacity at this point between hours 29 to 57. To estimate the maximum flood control level at the reservoir to ensure that point A will not be flooded, the flood control level of the reservoir is changed. Table 8.6 shows the results of the simulation of the system with different flood control storages in the reservoir, and the required flood control volume is about 5 million m^3.

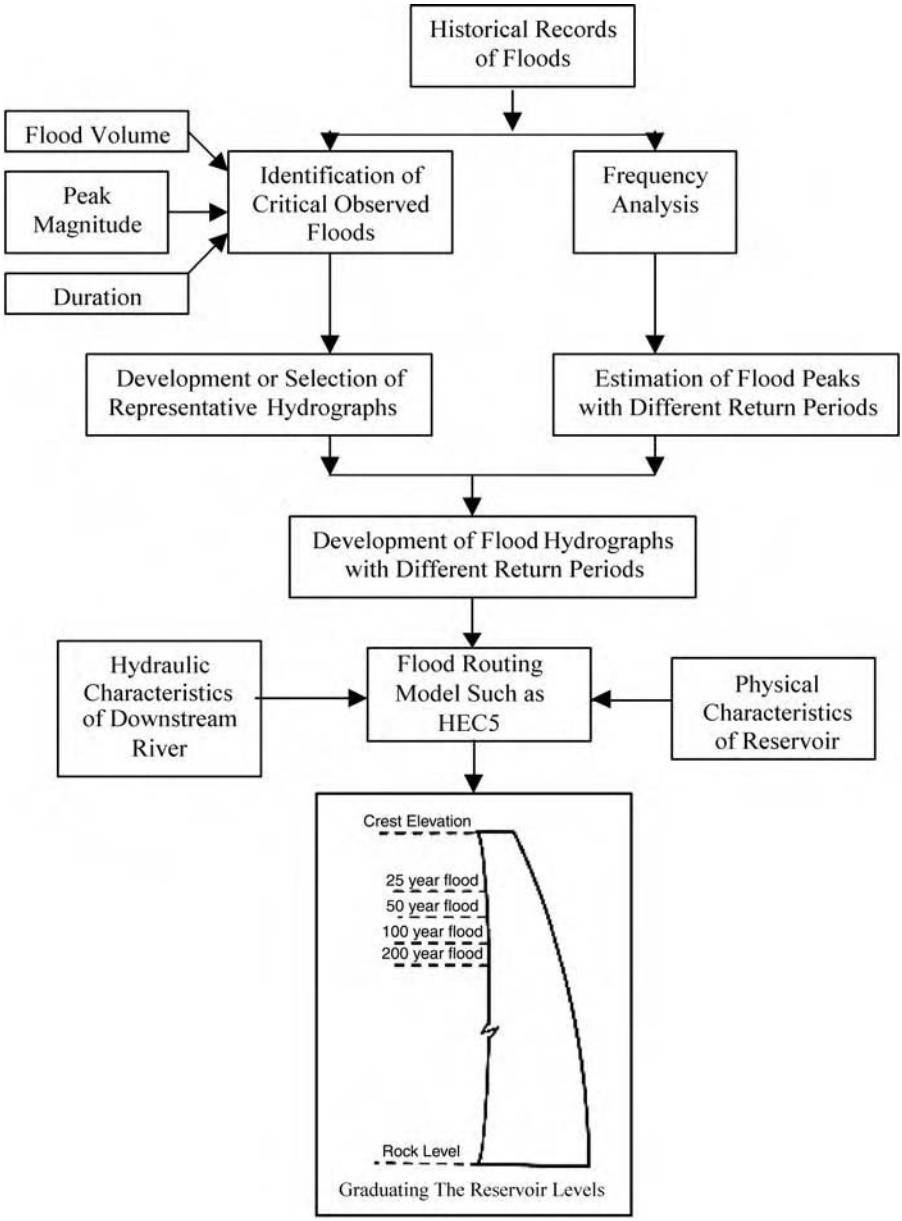


FIGURE 8.9 Flowchart for finding flood storage required for controlling floods with different return periods.

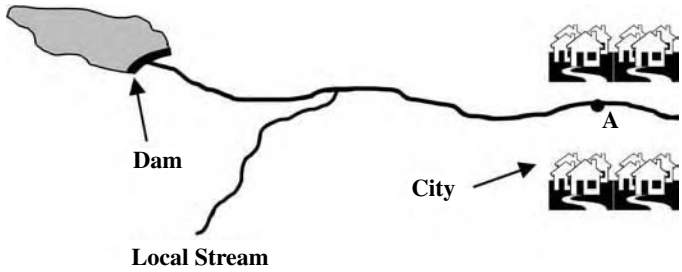


FIGURE 8.10 Schematic of the system explained in Example 8.4.

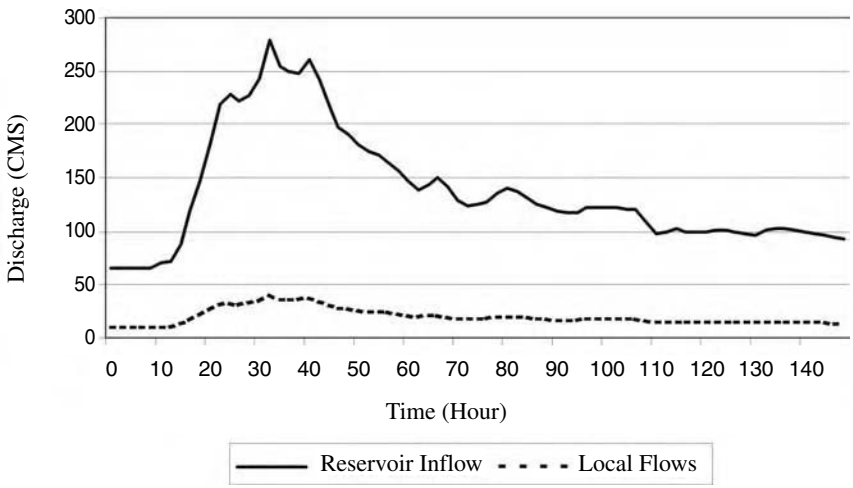


FIGURE 8.11 100-year hydrograph of reservoir inflow and local flows.

8.6 APPLICATION OF DETERMINISTIC DYNAMIC PROGRAMMING IN RESERVOIR OPERATION: DPR MODEL

The deterministic dynamic programming (DPR) model was developed by Karamouz et al. (1992) for optimization of multiple reservoir systems. The objective function of this model is expressed as:

$$\text{Min } Z = \sum_{t=1}^{T} Loss\left(\sum_{s=1}^{X} R_{s,t}\right) \tag{8.30}$$

where T is the time horizon, X is the total number of reservoirs in the system, and *Loss* is the cost of operation based on the ratio of supplied water ($R_{s,t}$) from reservoir s in month t to the monthly demand. As discussed earlier, this loss function is used

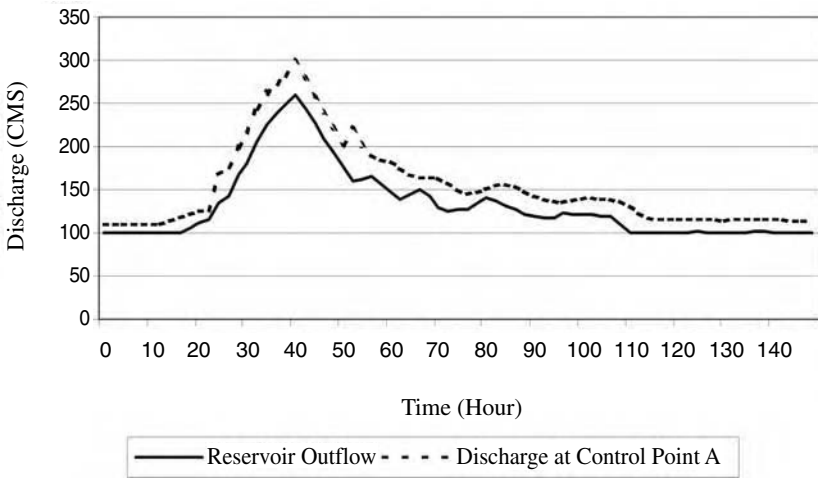


FIGURE 8.12 Reservoir outflow and discharge at control point A (the reservoir storage had been 67 million m^3 when the flood started).

TABLE 8.6
Results of Flood Routing in a River–Reservoir System (Example 8.4) for Various Flood Control Volumes

Reservoir Storage at Beginning of Flood (10^6 m^3)	Flood Control Volume (10^6 m^3)	Peak Discharge at Control Point A (m^3/sec)	Duration of Flooding Condition at A (hr)[a]
69	1	297.7	30
68	2	286.3	26
67	3	254.1	21
66	4	235.3	11
65	5	200.0	—

[a] Total number of hours that the discharge at the control point has been higher than 200 m^3/sec.

to prevent or minimize operation outside of a safe range. The continuity or mass balance of the contents of the reservoir, considering regulated and unregulated release (seepage and spill) from the beginning of the month to the next, is also included in the model:

$$S_{s,t+1} - S_{s,t} + R_{s,t} = I_{s,t} \tag{8.31}$$

$$R_{s,t} = R^R_{s,t} + R^U_{s,t} \tag{8.32}$$

where $R_{s,t}$ is release from reservoir s in month t, and I_t is inflow volume to the reservoir in month t. $R^R_{s,t}$ and $R^U_{s,t}$ are regulated and unregulated releases, respectively,

from reservoirs. Additional constraints on maximum and minimum allowable release and storage during any season for all sites can be stated as:

$$R_{s,t}^{\max} \geq R_{s,t} \geq R_{s,t}^{\min} \qquad t = 1, ..., T \quad s = 1, ..., X \tag{8.33}$$

$$S_{s,t}^{\max} \geq S_{s,t} \geq S_{s,t}^{\min} \qquad t = 1, ..., T \quad s = 1, ..., X \tag{8.34}$$

$$\left| S_{s,t+1} - S_{s,t} \right| \leq SC_s \qquad t = 1, ..., T \quad s = 1, ..., X \tag{8.35}$$

where SC_s is the maximum allowable change in storage of reservoir s within each month considering dam stability and safety conditions of the reservoir. This separable mathematical program is solvable as a discrete dynamic program. The recursion relation when Eq. (8.30) is used can be written as:

$$\begin{aligned} f_{t+1}(S_{1,t+1}, ..., S_{X,t+1}) = \min \Bigg(& Loss\left(\sum_{s=1}^{X} R_{st} \right) + f_t\left(S_{1,t}, ..., S_{X,t}\right) \Bigg) \\ & S_{1,t} \in \Omega_{1,t} \\ & \cdots \\ & \cdots \\ & S_{X,t} \in \Omega_{X,t} \end{aligned} \tag{8.36}$$

The initial conditions are:

$$f_1(S_{1,1}, ..., S_{X,1}) = 0 \qquad S_{1,1} \in \Omega_{1,1}, ..., S_{X,1} \in \Omega_{X,1} \tag{8.37}$$

where $f_t(S_{1,t}, \ldots, S_{X,t})$ is the total minimum losses of operation from the beginning of month 1 to the beginning of month t, when the storage volume at the beginning of month t is $S_{1,t}$ at site 1, …, $S_{X,t}$ at site X. $\Omega_{s,t}$ is the set of discrete storage volumes that will be considered for the beginning of month t at site s.

The optimal release ($\hat{R}_{s,m}$), optimal storage ($\hat{S}_{s,m}$), and inflows for each site can be regressed to estimate the general operating rules for different months in a simple form:

$$\hat{R}_{s,m} = a_{s,m} I_{s,m} + b_{s,m} \hat{S}_{s,m} + c_{s,m} \qquad m = 1, ..., 12 \quad s = 1, ..., X \tag{8.38}$$

where $a_{s,m}$, $b_{s,m}$, and $c_{s,m}$ are regression coefficients that should be estimated in each reservoir s for each month m. Initially, no upper bound on releases exists, and the lower bound on release is zero. The dynamic program is solved and the optimal policies are used in a multiple regression to find the general operating rules. These rules are then used to simulate the operation of reservoirs over a long time horizon. Also, the losses associated with simulated operation are determined. The procedure

concludes the first iteration of the model. For the second iteration, the allowable release in the dynamic program is limited to some percentage (bound) of the previously defined general operating rule, $\hat{R}_{s,t}$:

$$(1 - bound)\hat{R}_{s,t} \le R_{s,t} \le (1 + bound)\hat{R}_{s,t} \qquad s = 1, \ldots, X \tag{8.39}$$

Anytime the refined rule or best rule for a specific value of the bound is found, the algorithm can be further restricted by reducing the bound value, and the dynamic programming, regression analysis, and simulation procedure will continue. If the bound is very small (0.01), then the algorithm is stopped because further restrictions on the release in the dynamic program will probably not result in better operating rules.

8.7 STOCHASTIC RESERVOIR OPERATION MODELING

So far, the deterministic modeling approach has been discussed, where the state at the next stage is completely determined by the state and policy decision at the current stage. In other words, in the deterministic approach, it is assumed that what has happened in the past (e.g., historical inflows to the reservoirs) will be fully representative of what will happen in the future.

Considering the uncertainties inherent in the prediction of hydrologic, economic, and other factors affecting the performance of water resources systems, deterministic planning models may result in under- or overestimated results; systems benefits are generally overestimated, while systems costs are usually underestimated. Two types of uncertainty may affect reservoir operation models:

- *Uncertainty in objective function.* Models are subject to uncertainty in economic parameters and imprecise knowledge about future benefits and costs resulting from applying different policies. Such uncertainty can often be handled by substitution of the expected value for the uncertain objective value. Use of the expected value in the objective function usually results in satisfactory policies when the future states of the system are not too extreme (Loucks et al., 1981).
- *Uncertainty in constraints.* If the uncertainty is small (small variation in constraint variables), substitution of an expected value might be useful. When only the right-hand side of one or more inequality constraints is random, chance constraint programming can be used, as explained in the next section.

Two approaches have been developed in order to incorporate the stochastic nature of streamflows in the reservoir operation models:

- Incorporate the stochastic nature of streamflows implicitly in the reservoir operation models.
- Generate several synthetic streamflow records as inputs to deterministic models.

In the discussion that follows, linear and dynamic programming reservoir operation models in which the stochastic nature of streamflows is implicitly incorporated are explained.

8.7.1 Stochastic Linear Programming

As mentioned in Chapter 3, two approaches are considered in stochastic modeling of reservoir operation:

- Stochastic programming
- Chance constraint programming

The primary distinction between these types is that stochastic programming requires all constraints to hold with probability 1 and the objective function has probabilistic terms, whereas in chance constraint programming a small probability of violating any functional constraint is permitted but the objective function is usually considered to be deterministic. The general approach for dealing with the second type is to reformulate constraints as new equivalent linear programming problems for which the certainty assumption is satisfied.

In the first approach, the random variable (such as reservoir inflow) is usually represented by its discrete probability distribution and all the constraints associated with this random variable are repeated for values of the discretized variables. This results in stochastic programming models becoming computationally large (more details about stochastic optimization models are presented in Chapter 3). Chance constraint programming can be a suitable alternative approach, as explained in more detail in the next section.

8.7.1.1 Chance Constraint Reservoir Operation Model

Assume that reservoir storage at the end of each month (S_{t+1}) is a decision variable (d_t) that should be determined:

$$S_{t+1} = d_t \qquad (t = 1, \ldots, n) \tag{8.40}$$

The mass balance constraint would be as follows:

$$R_t = I_t - d_t + d_{t-1} \qquad (t = 1, \ldots, n) \tag{8.41}$$

Other constraints of the reservoir operation optimization model are the same as the deterministic model except for an additional constraint that represents the required water supply reliability:

$$\text{Prob}\left[R_t \geq D_t\right] \geq \lambda_t \qquad (t = 1, \ldots, n) \tag{8.42}$$

This equation shows that the probability of releases being greater than the demands (D_t) in the planning horizon should be greater than the reliability level (λ_t). This equation can be rewritten as:

$$\text{Prob}\left[I_t - d_t + d_{t-1} \geq D_t\right] \geq \lambda_t \qquad (t = 1, \ldots, n) \tag{8.43}$$

Each complementary probability ($1 - \lambda_t$) represents the allowable risk that the random variables will take on values such that:

$$I_t - d_t + d_{t-1} < D_t \tag{8.44}$$

Then, the deterministic equivalent of the probabilistic constrain can be written as:

$$D_t - d_{t-1} + d_t \leq i_t^{1-\lambda_t} \tag{8.45}$$

where $i_t^{1-\lambda_t}$ is the inverse of the cumulative distribution function of reservoir inflow at level λ_t such that:

$$P\left(I_t \leq i_t^{1-\lambda_t}\right) = 1 - \lambda_t \tag{8.46}$$

Example 8.5

Consider the following chance constraint problem assuming that $\tilde{b}_1$, $\tilde{b}_2$, and $\tilde{b}_3$ are normal variables with known mean and variances. Is X = (5, 10, 15) a feasible solution?

$$\text{Minimize} \quad Z = 10x_1 + 20x_2 + 30x_3$$

$$\text{Subject to:} \quad \text{Prob}\left[3x_1 + 2x_2 + x_3 \geq \tilde{b}_1\right] \geq 0.96$$

$$\text{Prob}\left[2x_1 + 4x_2 + 2x_3 \geq \tilde{b}_2\right] \geq 0.98$$

$$\text{Prob}\left[x_1 + 3x_2 + 5x_3 \geq \tilde{b}_3\right] \geq 0.99$$

$$\tilde{b}_1 = N(45, 4), \tilde{b}_2 = N(70, 9), \tilde{b}_3 = N(90, 25)$$

The deterministic equivalent of chance constraints can be computed as follows:

$$\begin{cases} 3x_1 + 2x_2 + x_3 \geq 45 + 2(1.75) = 48.5 \\ 2x_1 + 4x_2 + 2x_3 \geq 70 + 3(2.054) = 76.16 \\ x_1 + 3x_2 + 5x_3 \geq 90 + 5(2.327) = 101.64 \end{cases}$$

where the value in the parentheses are the z value from standard normal table for 0.96, 0.98, and 0.98 probability levels.

By substituting $X = (5, 10, 15)$ in the above deterministic equivalent of chance constraints, it can be seen that this point is a feasible solution:

$$\begin{cases} 3\times 5+2\times 10+15=50\geq 48.5 \\ 2\times 5+4\times 10+2\times 15=80\geq 76.16 \\ 5+3\times 10+5\times 15=110\geq 101.64 \end{cases}$$

8.7.2 Stochastic Dynamic Programming Model

Stochastic dynamic programming (SDP) differs from deterministic dynamic programming in that the state at the next stage is not completely determined by the state and policy decision at the current stage; rather, a probability distribution is used to determine what the next state will be. Figure 8.13 shows the basic structure of stochastic dynamic programming for reservoir operation. As can be seen in this figure, the state of the system in stage t can be represented by reservoir storage. Assuming a specific value for the decision variable (reservoir release) results in different states in stage $t + 1$ with probabilities that can be estimated based on associated probabilities with reservoir inflows in that time step.

The recursive function of a backward-moving stochastic dynamic programming algorithm can be written as follows:

$$f_t\left[S_t^c(k), I_t^c(l)\right]=\underset{S_{t+1}^c\in\Omega}{\text{Min}}\left[C_t\left[S_t^c(k), R_t\right]+\sum_{ll}P_{l,ll}^t f_{t+1}^*\left[S_{t+1}^c(kk), R_{t+1}\right]\right] \tag{8.47}$$

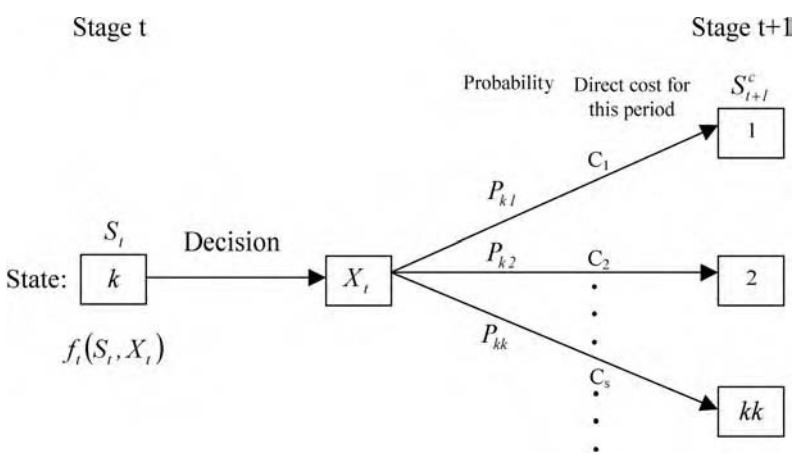


FIGURE 8.13 The basic structure for stochastic dynamic programming.

where:

Ω is the set of possible storage characteristics.
$S_t^c(k)$ is the reservoir storage characteristic value in state k in time t.
$I_t^c(l)$ is the reservoir inflow characteristic value in class interval l in time t.
$f_t\left[S_t^c(k), I_t^c(l)\right]$ is the minimum value of expected cumulative losses of reservoir operation from the present period to the end of the planning horizon, provided that the kth and lth scenario occurs in time t.
$f_{t+1}^*\left[S_{t+1}^c(kk), I_{t+1}^c(ll)\right]$ is the minimum value of expected losses during periods T to $t + 1$ associated with the optimal solution obtained when the kkth and llth scenario occurs in time $t + 1$.
$C_t\left[S_t^c(k), I_t^c(l), S_{t+1}^c(kk)\right]$ is the value of losses during period t, associated with the release R_t during the same time period, where $R_t = S_t^c(k) + I_t^c(l) - S_{t+1}^c(kk)$.
$P_{l,ll}^t$ is the probability of inflow in class interval l in period t, when the inflow in period $t + 1$ is in class interval ll.

The recursive function shown in Eq. (8.47) can be used in a procedure that moves backward. It is based on the principle that, no matter in what state of a stage, to find the best policy we should proceed from that state and stage in an optimal manner. Similar formulations can be developed for forward-moving procedures. This principle, first defined by Bellman (1957), is referred to as the *Bellman optimality principle* (Loucks et al., 1981).

The conditions for obtaining the steady-state optimal policy of the SDP model are:

- The expected value of the recursive function does not change from one period to the next (stationary solution).
- The optimal policies do not change.

For example, in reservoir operation, the optimal solution is reached when the optimal releases associated with each initial storage are the same as the corresponding releases and storage in the previous time step. Inflow and storage discretization significantly affect the performance of stochastic dynamic programming models for reservoir operation optimization.

8.7.2.1 Reservoir Storage Discretization

Due to the nature of dynamic programming, the storage volume must be discretized into representative storage zones, with a corresponding number of characteristic storage levels (*SDN*). In the classical scheme, Z is simply defined as the number of zones equal to *SDN* and the storage increment is:

$$\Delta S = \frac{Cap}{SDN} \tag{8.48}$$

where *Cap* is the total permissible storage capacity of the reservoir. For each state k at time t, the characteristic storage states can be calculated as:

$$S_t^c(k) = \frac{\Delta S}{2} + (k-1)\Delta S = \frac{2k-1}{2}\Delta S \qquad \forall k \in [1, ..., SDN] \tag{8.49}$$

In this approach, the zero and full storage levels are not included as separate characteristic storage values. Savaranskiy (1940) developed another scheme in which each Z zone is regarded as a class interval, and the corresponding storage state is defined as the center point of this interval. Empty and full reservoir storages are also defined as separate states with zero class intervals:

$$Z = SDN - 1 \qquad \text{and} \qquad \Delta S = \frac{Cap}{SDN - 2} \tag{8.50}$$

So, the characteristic values can be estimated as follows:

$$S_t^c(k) = \frac{\Delta S}{2} + (k-2)\Delta S = \frac{2k-3}{2}\Delta S \qquad k = 2, \cdots, Z \tag{8.51}$$

Moran (1945) presented another method in which the characteristic storage values are defined on the boundaries between the zones:

$$Z = SDN - 1 \qquad \text{and} \qquad \Delta S = \frac{Cap}{SDN - 1}$$

and storage stages can be calculated as

$$S_t^c(k) = (k-1)\Delta S \qquad k = 1, \cdots, Z \tag{8.52}$$

EXAMPLE 8.6

The capacity of a reservoir is 16.8 million m^3. Find the storage zones using the discretization schemes explained previously.

Solution: Figure 8.14 shows storage zones obtained from discretization schemes for $SDN = 8$.

8.7.2.2 Reservoir Inflow Discretization

The governing idea for flow discretization consists of dividing the entire flow range into a specific number of class intervals (*DN*) and calculating a characteristic flow value ($I_t^c(l)$) for each interval. This value represents all the flows of that class

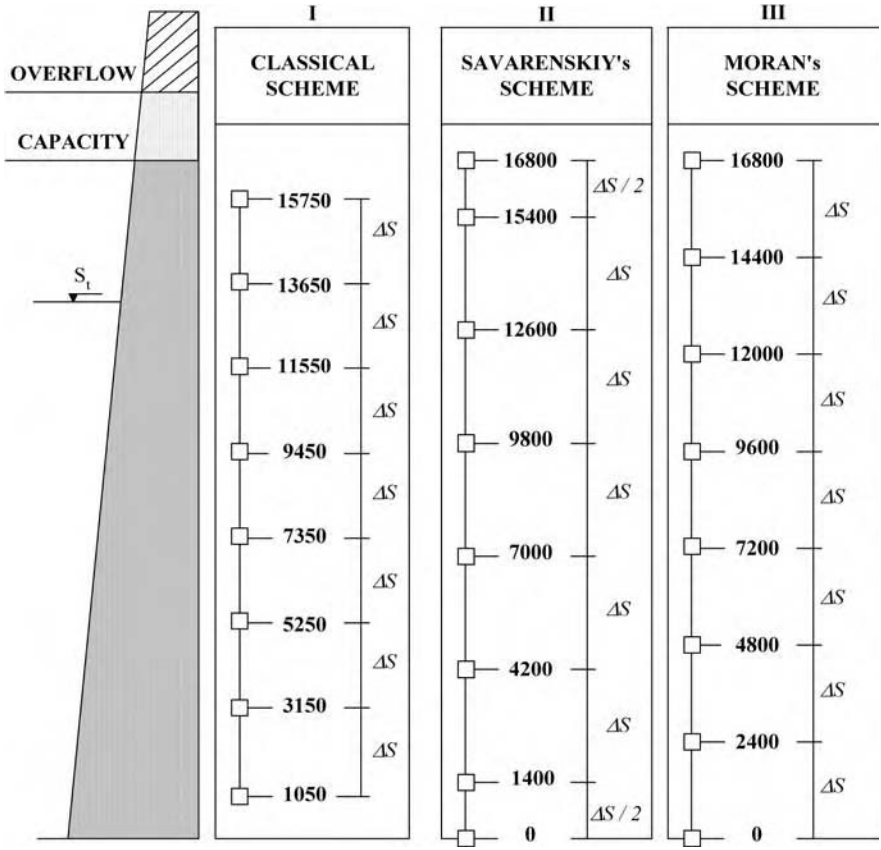


FIGURE 8.14 Discrete representation of storage in 1000 m^3 (for capacity equal to 16.8 million m^3 and *SND* = 8). (From Karamouz and Vasiliadis, 1992.)

interval. The classical scheme that has been used by many researchers consists of dividing the entire flow range (ΔI), from I_{min} to I_{max}, into *DN* class intervals wit constant flow increment δI, where I_{min} and I_{max} are minimum and maximum inflows, respectively. In this scheme:

$$\delta I = \frac{\Delta I}{DN} \tag{8.53}$$

The characteristic flow ($I_t^c(l)$) for each class interval is defined as the mean of the lower and upper bounds of the corresponding class interval. In other words, the entire discretization model is based on the minimum and maximum values of each interval, and any changes of these values will result in completely different characteristic values.

An improved approach is to divide class interval into a number of subintervals and define a probability distribution of inflows occurring in each class interval using frequency analysis for each class subinterval. The characteristic value for (representative of) each interval is then defined as the expected value of inflows within that interval. This and the previous approach have some shortcomings. First, the variance of the characteristic flows could be very different from those of the original series. Furthermore, some intervals might have fewer data, which could result in a poor estimation of transition probabilities from these intervals to others.

To overcome these shortcomings, another method was developed by Karamouz and Vasiliadis (1992), where the intervals are unequally spaced. In this model, the amount of data ($N(i)$) for each class interval is made the same (so the intervals are equally probable) by adjusting the width of each class interval (see Figure 8.15), and the variance of data within each class interval is kept small by increasing the number of class intervals. If $N(i)$ is the number of data points in the ith class interval, $n(i, j)$ is the number of data points in the ith class interval and the jth subinterval (where $n(i, j) < N(i)$), $f^c(i, j)$ is the average of the lower and the upper bound of the jth subinterval, and $SI(i)$ is the number of subintervals in the ith class, then the characteristic value for each class interval, $F^c(i)$, is calculated using the following formula:

$$F^c(i) = \sum_{j=1}^{SI(i)} \left[\frac{n(i,j)}{N(i)} \cdot f^c(i,j) \right] \qquad \forall i \in [1, \cdots, DN] \qquad n(i,j) \in [1, \cdots, N(i)], \quad N(i) \in [1, \cdots, N] \tag{8.54}$$

An alternative expression for Eq. (8.54) is:

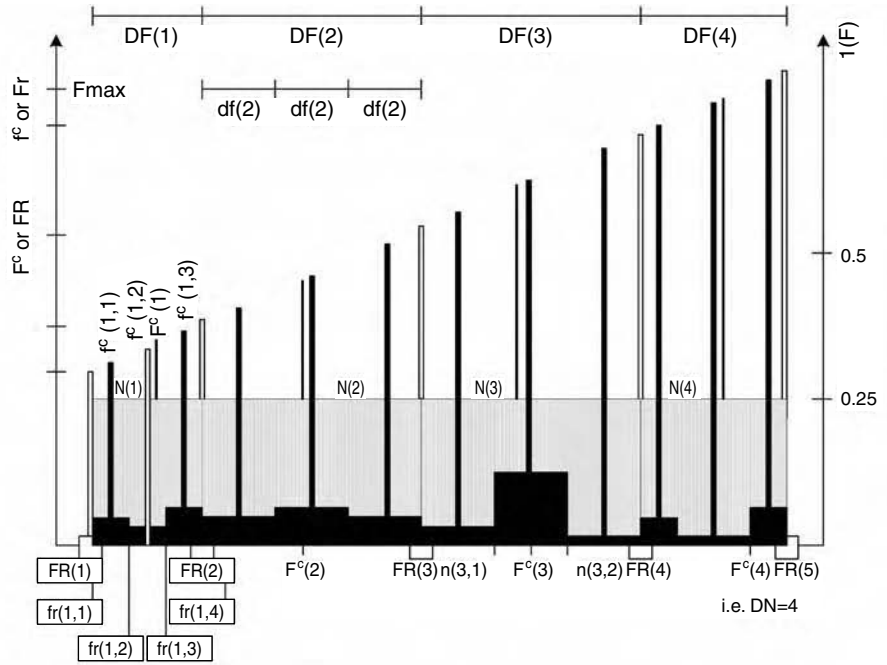
$$F^c(i) = \sum_{j=1}^{N(i)} \left[\frac{x(i,j)}{N(i)} \right] \qquad \forall i \in [1, \cdots, DN], N(i) \in [1, \cdots, N] \tag{8.55}$$

where $x(i, j)$ is the jth flow in the ith class interval. This method of selecting characteristic flows results in better transition matrices because an adequate number of flows is available for each interval, zero entries in the transition matrices are eliminated, and the value of each data point receives equal weight in the construction of the transition matrices. Furthermore, the characteristic flows preserve the mean and variance of the original time series, as long as the number of class intervals is sufficiently large.

Example 8.7

The monthly inflows to a reservoir have been:

$$q = \{5, 1, 6, 3, 4, 7, 8, 2, 10, 7\}$$



LEGEND

- Characteristic Flow
- Bounds of Class Intervals
- Characteristic Flow of Class Sub-Intervals
- Frequency Histogram
- Frequency Histogram within each Class Interval

EQUATIONS:

Fmn = FR(1) = fr(1,1)

Fmax = FR(DN+1) = fr(DN,dn+1)

N(1)=INT[N/DN] (Constant)

note: FR(i) is the [N(i) x(i-1)+1]th flow, after sorting

dn(i) = INT[1+3.3 xlog(N(i))] (Constant)

df(i) =[FR(i+1)- FR(i)] / dn(1) (Constant)

fr(i,j) = df(i)x(j-1)+F(i)

$f^c(i,j) = [rf(i,j) + fr(i,j+1)] / 2$

$$F^c(i) = \sum_{\text{all } j} \frac{n(i,j)}{N(i)} * f^c(i,j)$$

all i in [1,...,DN]
all j in [1,...,dn(i)]

FIGURE 8.15 Discrete representation of flow. (From Karamouz and Vasiliadis, 1992.)

The transition probabilities of inflows considering two intervals for $q \le 5$ and $q > 5$ are shown in the following matrix:

$$P\left(q_{t+1} \middle| q_t\right) = \begin{array}{c} \\ q \le 5 \\ q > 5 \end{array} \begin{array}{c} \begin{array}{cc} q \le 5 & q > 5 \end{array} \\ \begin{bmatrix} 0.4 & 0.6 \\ 0.5 & 0.5 \end{bmatrix} \end{array}$$

Consider the mass balance equation for the reservoir ($R_t = S_t - S_{t+1} + q_i$) and two storage intervals with the characteristic values $S_t = \{1, 2\}$. The cost of operation can be estimated as $f_t = R_t^2$. The characteristic volumes of the streamflows are considered to be 3 and 7. Find the optimal releases using the stochastic dynamic programming technique.

Solution: The release from the reservoir for each month can be estimated by considering different storage levels at the beginning and end of the month. The stochastic dynamic programming solution for finding the optimal releases is shown in Table 8.7. Using the backward-moving procedure in the first iteration, which is assumed to be the 10th time interval in the planning horizon, the optimal storage level at the beginning of the month and releases are estimated considering possible inflow and ending storage states. In the next iterations, the cumulative cost of operation is calculated considering the inflow transition probabilities.

For example, in the first row of results for the 9th time step, storage at the end of the month is considered to be 1; therefore, the cost of the best policies obtained in the 10th time step for starting storage equal to 1 should be added to the cost of operation in the 9th time step. The least costs are estimated as 4 and 36 when the starting storage is 1 and inflows in the 10th time step are 3 and 7, respectively. As can be seen for the second and third iterations, the optimal releases and storages obtained are the same and the expected cost in one time step remains constant; therefore, the stationary solution has been obtained.

8.7.3 BSDP and DDSP Models

The Bayesian Stochastic Dynamic Programming (BSDP) model was developed by Karamouz and Vasiliadis (1992). In this model, the release from a reservoir in each time period t depends on the state of the system, which is defined by the following state variables:

- Characteristic value of reservoir storage at the beginning of time period t ($S_t^c(k)$)
- Characteristic inflow to the reservoir during time period t ($I_t^c(l)$)
- Characteristic forecast for the next time period, $t + 1$ ($H_{t+1}^c(ii)$)

For every state $[S_t^c(k), I_t^c(l), H_{t+1}^c(ii)]$, the optimal release of R_t^* is obtained by solving the following DP recursion equation:

$$f_t\left[S_t^c(k), I_t^c(l), H_{t+1}^c(ii)\right] = \min\left\{B_t\left[S_t^c(k), I_t^c(l), S_{t+1}^c(kk)\right] + \alpha \sum_{ll=1}^{IDN}\left[\phi\left[I_{t+1}^c(ll)\middle|H_{t+1}^c(ii), I_t^c(l)\right]\right.\right.$$
$$\left.\left.\cdot \sum_{iv=1}^{FDN}\left[\pi\left[H_{t+2}^c(iv)\middle|I_{t+1}^c(ll), H_{t+1}^c(ii), I_t^c(l)\ \right]\cdot f_{t+1}^*\left[S_{t+1}^c(kk), I_{t+1}^c(ll), H_{t+2}^c(iv)\ \right]\right]\right]\right\} \tag{8.56}$$

where:

$B_t[S_t^c(k), I_t^c(l), S_{t+1}^c(kk)]$ is the loss during time period t.

$\phi_{t+1}(I_{t+1}^c(ll) \mid H_{t+1}^c(ii), I_t^c(l))$ is the posterior flow transition (mass function) from the forecast in the next month and the current actual flow to the next actual flow.

$\pi[H_{t+2}^c(iv) \mid I_{t+1}^c(ll), H_{t+1}^c(ii), I_t^c(l)]$ is the posterior forecast transition probability function from the next actual and predicted flow in time period $t + 1$ and the current actual flow to future predicted flow in time period $t + 2$.

In this model, the posterior probabilities are estimated considering the Bayesian process. The prior flow transition probabilities $\rho_{t+1}[I_{t+1}(l) \mid I_t(lm)]$ and likelihoods $\lambda_{t+1}[H_{t+1}(i) \mid I_{t+1}(l)]$ can be calculated from the historical and predicted inflow time series. Then, the posterior probabilities are estimated as follows:

$$\phi_{t+1}\left(I_{t+1}^c(ll)\middle|H_{t+1}^c(ii), I_t^c(l)\right) = \frac{\lambda_{t+1}\left[H_{t+1}(i)\middle|I_{t+1}(l)\right]\cdot\rho_{t+1}\left[I_{t+1}(l)\middle|I_t(lm)\right]}{\sum_{l=1}^{IDN}\lambda_{t+1}\left[H_{t+1}(i)\middle|I_{t+1}(l)\right]\cdot\rho_{t+1}\left[I_{t+1}(l)\middle|I_t(lm)\right]} \tag{8.57}$$

$$\pi\left[H_{t+2}^c(iv)\middle|I_{t+1}^c(ll), H_{t+1}^c(ii), I_t^c(l)\right] = \sum_{ll=1}^{IDN}\left\{\lambda_{t+1}\left[H_{t+2}(ii)\middle|I_{t+2}(ll)\right]\right.$$
$$\cdot\sum_{ii=1}^{FDN}\left[\phi_{t+1}\left[I_{t+2}(ll)\middle|H_{t+2}(ii), I_{t+1}(l)\right]\right.$$
$$\left.\left.\cdot\sum_{l=1}^{IDN}\left\{\left[\sum_{l=1}^{IDN}\lambda_{t+1}\left[H_{t+1}(i)\middle|I_{t+1}(l)\right]\cdot\rho_{t+1}\left[I_{t+1}(l)\middle|I_t(lm)\right]\right]\right.\right.\right.$$
$$\left.\left.\left.\cdot\phi_{t+1}\left(I_{t+1}^c(ll)\middle|H_{t+1}^c(ii), I_t^c(l)\right)\right\}\right] \tag{8.58}$$

TABLE 8.7
SDP Solution for Reservoir Operation Problem (Example 8.6)

t	S_t	q_t	S_{t+1}	q_{t+1}	R_t	f_t	$P(q_t \mid q_{t-1})$	$f_t(S_t, R_t)$	R_t^*	S_{t+1}^*
First Iteration										
10	1	3	1	3	3	9	0.4	9	2	2
				7			0.6			
			2	3	2	4	0.4	4		
				7			0.6			
		7	1	3	7	49	0.5	49	6	2
				7			0.5			
			2	3	6	36	0.5	36		
				7			0.5			
	2	3	1	3	4	16	0.4	16	3	2
				7			0.6			
			2	3	3	9	0.4	9		
				7			0.6			
		7	1	3	8	64	0.5	64	7	2
				7			0.5			
			2	3	7	49	0.5	49		
				7			0.5			

Second Iteration										
9	1	3	1	3	3	9	0.4	$9 + 0.4 \times 4 + 0.6 \times 36 = 32.2$	3	1
				7			0.6			
			2	3	2	4	0.4	$4 + 0.4 \times 9 + 0.6 \times 49 = 37$		
				7			0.6			
		7	1	3	7	49	0.5	$49 + 0.5 \times 4 + 0.5 \times 36 = 69$	6	2
				7			0.5			
			2	3	6	36	0.5	$36 + 0.5 \times 9 + 0.5 \times 49 = 65$		
				7			0.5			
	2	3	1	3	4	16	0.4	$16 + 0.4 \times 4 + 0.6 \times 36 = 39.2$	4	1
				7			0.6			
			2	3	3	9	0.4	$9 + 0.4 \times 9 + 0.6 \times 49 = 42$		
				7			0.6			
		7	1	3	8	64	0.5	$64 + 0.5 \times 4 + 0.5 \times 36 = 84$	7	2
				7			0.5			
			2	3	7	49	0.5	$49 + 0.5 \times 9 + 0.5 \times 49 = 78$		
				7			0.5			

(continued)

TABLE 8.7 (CONTINUED)
SDP Solution for Reservoir Operation Problem (Example 8.6)

t	S_t	q_t	S_{t+1}	q_{t+1}	R_t	f_t	$P(q_t \mid q_{t-1})$	$f_t(S_t, R_t)$	R_t^*	S_{t+1}^*
Third Iteration										
8	1	3	1	3	3	9	0.4	$9 + 0.4 \times 32.2 + 0.6 \times 65 = 60.88$	3	1
				7			0.6			
			2	3	2	4	0.4	$4 + 0.4 \times 39.2 + 0.6 \times 78 = 66.48$		
				7			0.6			
		7	1	3	7	49	0.5	$49 + 0.5 \times 32.2 + 0.5 \times 65 = 97.6$	6	2
				7			0.5			
			2	3	6	36	0.5	$36 + 0.5 \times 39.2 + 0.5 \times 78 = 94.6$		
				7			0.5			
	2	3	1	3	4	16	0.4	$16 + 0.4 \times 32.2 + 0.6 \times 65 = 67.88$	4	1
				7			0.6			
			2	3	3	9	0.4	$9 + 0.4 \times 39.2 + 0.6 \times 78 = 71.4$		
				7			0.6			
		7	1	3	8	64	0.5	$64 + 0.5 \times 32.2 + 0.5 \times 65 = 112.6$	7	2
				7			0.5			
			2	3	7	49	0.5	$49 + 0.5 \times 39.2 + 0.5 \times 78 = 107.6$		
				7			0.5			

The Demand-driven Stochastic Dynamic Programming (DDSP) model was developed by Vasiliadis and Karamouz (1994) to generate operating policies. It accounts not only for the natural and forecast uncertainties of the system (as does BSDP) but also for the seasonal (monthly) variation of various parameters (such as inflow and demand) in the system. The major characteristics of the DDSP model are as follows:

- Flow transition probabilities associated with seasonal variations are continuously updated on a monthly basis using Bayesian decision theory (BDT).
- The model is constructed to accept variable and uncertain demands.
- Each iteration is considered to represent 12 months of a year at a time, thus the computational efforts are reduced.
- The cost function includes release and expected storage.

A flow discretization scheme developed by Karamouz and Vasiliadis (1992) is used in this model. With this scheme, the characteristics flow values preserve the mean and standard deviation of the flow time series. In DDSP, the decision variable (release) depends on the state of the system, which is defined by the following variables:

M, number of month ($M = 1, \ldots, 12$)
D_M, corresponding demand in month M
S_M, reservoir storage at the beginning of month M ($S_M = 0, \ldots, Cap$)
I_M, inflow to the reservoir during month M ($I_M = I_1, \ldots, I_{IDN}$)
H_{M+1}, forecast of flow for the next month ($M + 1$), where $H_{M+1} = H_1, \ldots, H_{FDN}$

The recursive function at year Y and month M should be written in terms of characteristic values of the above variables:

$$f_{Y,M}(D_M^c, S_M^c, I_M^c, H_{M+1}^c) = \text{Min}\{E(D_{M+1}^c).CR_M + CS_{M+2} + (E(I_{M+1}^c \mid H_{M+1}^c, I_M^c) .\{E(H_{M+2}^c \mid I_{M+1}^c).[f^*_{Y,M+1}(D_{M+1}^c, S_{M+1}^c, I_{M+1}^c, H_{M+2}^c)]\})]\} \tag{8.59}$$

where:

Y is the index of years/iterations.
CR_M is the cost related to release during month M.
CS_{M+2} is the cost related to expected ending storage in month $M + 2$.
E is the expected value.
D_M^c is the characteristic value of water demand in month M.
S_M^c is the characteristic value of reservoir storage at the beginning of month M.
I_M^c is the characteristic value of inflow to the reservoir during month M.
H_M^c is the characteristic value of inflow forecast for month M.

In Eq. (8.59), where the demand uncertainty is incorporated in the expected value of demands ($E(D_{M+1})$), it is assumed that inflow and demand are independent stochastic variables. However, if the water supply (inflow) and demand are site specific, the expectation of $E(D_{M+1} \mid I_M)$ can be used in this equation. Formulation of the loss function and the optimal operation policies for this model are explained in the next sections.

8.8 MULTIPLE RESERVOIR OPERATION MODELING

In this section, the single reservoir operation model is extended to a system of cascade and/or parallel reservoirs. The basic issue examined here is how much water can be supplied to different demand points from which sources. As an example, the framework of an optimization model for a system of reservoirs consisting of both cascade and parallel reservoirs, taking into consideration priority-based heuristics for water allocation, is also presented in the following sections.

8.8.1 Cascade Reservoirs

Cascade systems have more than one reservoir on the same river (see Figure 8.16). It is assumed that demand points are located along the river downstream of the reservoirs. Also, local rivers join the main river at various reaches. The optimization model for this system is as follows:

$$\text{Minimize} \quad z = \sum_{t=1}^{n} \sum_{i=1}^{NR} \sum_{j=1}^{m_i} loss_{i,j,t}\left(RA_{i,j,t}, D_{i,j,t}, S_{i,t}\right) \tag{8.60}$$

$$\text{Subject to:} \quad S_{1,t+1} = S_{1,t} + I_{1,t} - R_{1,t} - E_{1,t} - L_{1,t} \qquad (t = 1,\ldots,n) \tag{8.61}$$

$$S_{i,t+1} = S_{i,t} + \sum_{j=1}^{m_i}\left(R_{i-1,t} - RA_{i-1,j,t}\right) + q_{i-1,t} - R_{i,t} - E_{i,t} - L_{i,t}, \quad (t = 1,\ldots n) \quad (i = 2,\ldots,NR) \tag{8.62}$$

$$I_{i,t} = R_{i-1,t} - \sum_{j=1}^{m_i} RA_{i,j,t} + q_{i-1} \qquad (t = 1,\ldots,n)(i = 2,\ldots,NR) \tag{8.63}$$

$$S_{i,\min} \le S_{i,t} \le Cap_i \qquad (t = 1,\ldots,n)(i = 1,\ldots,NR) \tag{8.64}$$

$$0 \le R_{i,t} \le R_{i,\max,t} \qquad (t = 1,\ldots,n)(i = 1,\ldots,NR) \tag{8.65}$$

$$S_{i,t}, E_{i,t}, L_{i,t}, R_{i,t} \ge 0 \qquad (t = 1,\ldots,n)(i = 1,\ldots,NR) \tag{8.66}$$

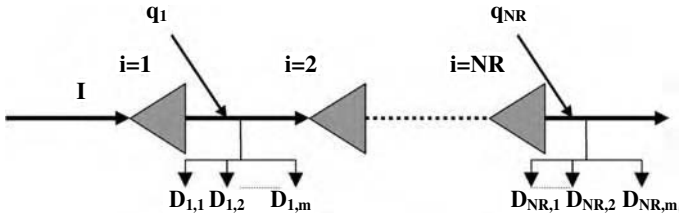


FIGURE 8.16 Cascade reservoir system.

where:

$RA_{i,j,t}$ is the water allocated to demand j from reservoir i in month t (volume of water allocated to the demand j minus volume of return flows in reach i).
$R_{i,t}$ is the release from reservoir i in month t.
$R_{i,max,t}$ is the maximum allowable release from reservoir i in month t.
$q_{i,t}$ is the local flow downstream of reservoir i in month t.
$E_{i,t}$ is the volume of water evaporated from reservoir i in month t.
$L_{i,t}$ is the volume of water leakage from reservoir i in month t.
$S_{i,t}$ is the volume of water storage in reservoir i at the beginning of month t.
NR is the number of river reaches.
$D_{i,j,t}$ is the water demand j downstream of reservoir i in month t.
Cap_i is the total capacity of reservoir i.
$S_{i,min}$ is the minimum allowable storage in reservoir i.
$I_{i,t}$ is the inflow to reservoir i in month t.
m_i is the number of demands in reach i.

Using this model, the contribution of each reservoir to supplying the different demands can be determined.

8.8.2 Parallel Reservoirs

Parallel reservoirs have a junction point with other rivers in the system. It is assumed that demand points are located downstream of each reservoir and after the junction point. Reservoirs in parallel require special modeling when demand points downstream of the junction point should be supplied from upstream reservoirs; the state of each reservoir for supplying direct demands must then be considered. Formulation of the optimization problem can be as follows:

$$\text{Minimize} \quad z = \sum_{t=1}^{n} \sum_{i=1}^{NR+1} \sum_{j=1}^{m_i} loss_{i,j,t}\left(RA_{i,j,t}, D_{i,j,t}, S_{i,t}\right) \tag{8.67}$$

$$\text{Subject to:} \quad S_{i+1,t+1} = S_{i,t} + I_{i,t} - R_{i,t} - E_{i,t} - L_{i,t} \qquad (t = 1,\ldots,n)(i = 1,\ldots,NR) \tag{8.68}$$

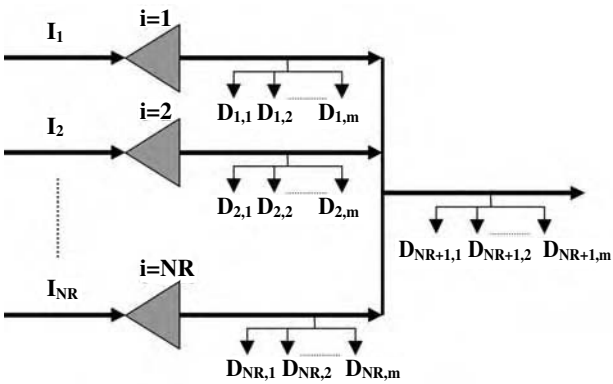


FIGURE 8.17 Parallel reservoir system.

$$I_{NR+1,t} = \sum_{i=1}^{NR} R_{i,t} - \sum_{j=1}^{m_i} RA_{i,j,t} \qquad (t = 1,\ldots,n) \tag{8.69}$$

$$S_{i,\min} \le S_{i,t} \le Cap_i \qquad (t = 1,\ldots,n)(i = 1,\ldots,NR) \tag{8.70}$$

$$0 \le R_{i,t} \le R_{i,\max,t} \qquad (t = 1,\ldots,n)(i = 1,\ldots,NR) \tag{8.71}$$

$$S_{i,t}, E_{i,t}, L_{i,t}, R_{i,t} \ge 0 \qquad (t = 1,\ldots,n)(i = 1,\ldots,NR) \tag{8.72}$$

By solving the above formulation, the contribution of each reservoir toward supplying the demand points along the $NR + 1$ reach is determined.

Example 8.8

Consider a system of three reservoirs as shown in Figure 8.18. In this system, two parallel reservoirs are in series with a single reservoir; demand points are considered downstream of each reservoir. As shown in Figure 8.18, agricultural lands are downstream of dam 1. Also, towns A and B are located downstream of dam 2 and dam 3; navigation and the instream flow rights exist downstream and upstream of dam 3, respectively. A third river joins the releases of dams 1 and 2 in the junction point of the two downstream rivers. Formulate the optimization model for finding optimal outflows from these reservoirs.

Solution: The following formulation may be used for optimization of this multiple reservoir system:

$$\text{Minimize} \quad z = \sum_{t=1}^{n} \sum_{i=1}^{3} loss_{i,t}\left(RA_{i,t}, D_{i,t}, S_{i,t}\right) \tag{8.73}$$

$$\text{Subject to:} \quad S_{i+1,t+1} = S_{i,t} + I_{i,t} - R_{i,t} - E_{i,t} - L_{i,t} \qquad (t = 1,\ldots,n)\ (i = 1,2) \tag{8.74}$$

$$S_{3,t+1} = S_{3,t} + I_{3,t} + \sum_{i=1}^{2} (R_{i,t} - RA_{i,t}) - E_{3,t} - L_{3,t} \qquad (t = 1,\ldots,n) \tag{8.75}$$

$$S_{i,\min} \le S_{i,t} \le Cap_i \qquad (t = 1,\ldots,n)(i = 1,2,3) \tag{8.76}$$

$$0 \le R_{i,t} \le R_{i,\max,t} \qquad (t = 1,\ldots,n)(i = 1,2,3) \tag{8.77}$$

$$\sum_{i=1}^{2} \left(R_{i,t} - RA_{i,t}\right) + I_{3,t} \ge Q_{u,m} \quad (t = 1,\ldots,n)\ \left(m = t - 12\left[(t-1)/12\right]\right) \tag{8.78}$$

$$(R_3 - RA_{i,3}) \ge Q_{d,m} \quad \forall t \left(m = t - 12\left[(t-1)/12\right]\right) \tag{8.79}$$

$$S_{i,t}, E_{i,t}, L_{i,t}, R_{i,t} \ge 0 \quad (t = 1,\ldots,n)(t = 1,\ldots,3) \tag{8.80}$$

where:

Cap_i is the total capacity of dam i.
$D_{1,m}$ is the water demand of agricultural lands downstream of dam 1 in month m.
$D_{2,m}$ is the water demand of town A downstream of dam 2 in month m.
$D_{3,m}$ is the water demand of town B downstream of dam 3 in month m.
$Q_{u,m}$ is the minimum instream water required upstream of dam 3 in month m.
$Q_{d,m}$ is the minimum water required for navigation downstream of dam 3 in month m.
$RA_{i,t}$ is the water allocated to demand downstream of dam i in month t.

8.9 RESERVOIR OPERATING RULES

Real-time operation of river–reservoir systems requires specific operating rules. These rules are guides for water conservation and release policies prepared for reservoir operators. The several types of rules range from very simple and static to dynamic for considering the varying states of inflow and physical characteristics of a reservoir in each time period. One of the simplest rules for reservoir operation is the *rule curve*, which specifies the target storage at the end of each month

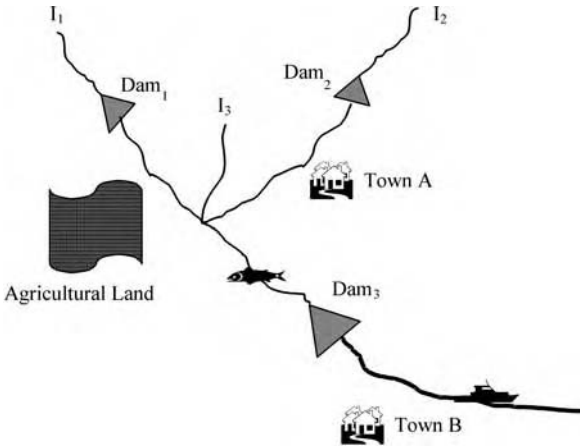


FIGURE 8.18 Example of a multiple reservoir system.

(Figure 8.19). Rule curves are static forms of operating policies that do not get any feedback from reservoir storage and current hydrologic situations such as predicted inflows to reservoirs in the following months. Rule curves are not very efficient policies, particularly when inflows and demands are highly varied, but they have been widely used because of their simplicity.

Application of deterministic and stochastic optimization models for reservoir operation has led investigators to define nonstatic types of operating rules. Karamouz et al. (1992) have used the following form of operating rule in the DPR model for optimization of multiple reservoir systems:

$$R_t = a_m I_t + b_m S_t + c_m \quad \left(m = t - 12\left[(t-1)/12\right]\right) \tag{8.81}$$

where a_m, b_m, and c_m are rule constants in month m. In DPR models, the optimal releases, optimal storages, and historical inflows are regressed to estimate the above operating rule. More complex nonlinear forms of operating rules are also tested by different researchers. Results of these studies have shown that the simple linear rules are as good as or sometimes better than nonlinear rules (Bhaskar and Whitlatch, 1980). This type of operating policy takes into account the state of the system in each month and has shown a better reliability in satisfying different objectives of the reservoir operation.

Another type of operation policy is produced by stochastic dynamic programming models. In these policies, the decision variable (release) depends on variables representing the state of the system for each month, including:

- Inflow to the reservoir
- Water demand
- Storage at the beginning of a month
- Inflow forecast for the next month

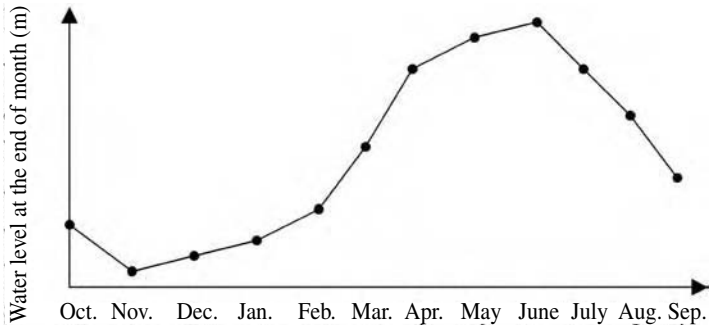


FIGURE 8.19 Typical reservoir rule curve.

TABLE 8.8
DDSP Model Optimal Operation Policies

No. of Month	Demand in Month t (10^6 m^3)	Storage at First of Month t (10^6 m^3)	Inflow Forecast in Month t (10^6 m^3)	Inflow Forecast in Month $t+1$ (10^6 m^3)	Release in Month t (10^6 m^3)
1	20	150	35	35	15
1	20	150	35	30	15
1	20	150	35	25	10
1	20	150	30	35	15
1	20	150	30	30	10
1	20	150	30	25	10
1	20	150	25	35	15
1	20	150	25	30	15
1	20	150	25	25	5

Therefore, the operation policies in these models are like the contents of a table in which various combinations of characteristic values for a state variable and the optimal release for that selection are presented in each row. Table 8.8 shows a part of the optimal operating policies derived from the DDSP model as developed by Vasiliadis and Karamouz (1994). The DDSP model is a demand-driven stochastic dynamic programming model that takes into account all of the above state variables.

Neural networks and fuzzy systems have been widely used for reservoir operation modeling. These techniques estimate input/output functions, and all are trainable dynamic systems. Unlike statistical estimators, they estimate a function without a mathematical model of how outputs depend on inputs (Hasebe and Nagayama, 1996). Application of these techniques for defining reservoir operation policies has been considered by many investigators, including Raman and Chandramouli (1996) and Shrestha et al. (1996).

Artificial neural networks consist of numerous, simple processing units or "neurons" that can be globally programmed for computation. In reservoir operation

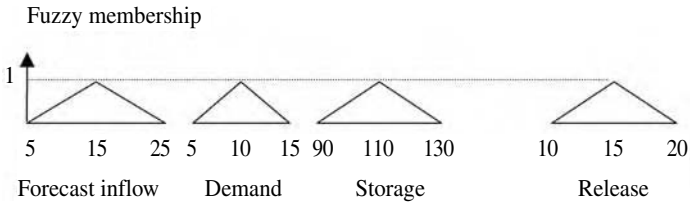


FIGURE 8.20 An example of fuzzy reservoir operating rules.

modeling, historical records of inflow to the reservoirs, demands, and optimal values of release and storage in the reservoir obtained from any optimization algorithm can be used for training a neural network system. This trained neural network can be used as an operating policy in real-time operation.

In a similar case, these inputs can be used for defining fuzzy rules. The following steps should be taken to develop fuzzy rules:

1. Divide input/output space into fuzzy regions.
2. Generate fuzzy rules from the given data.
3. Calculate the degree of fulfillment for each rule.
4. Assign a weight to each rule.

Various types of fuzzy membership functions, such as triangular or trapezoidal shapes, have been used by investigators (Figure 8.20). Selection of the proper shape and domain of the membership functions can be done based on the variability of each input variable and the sensitivity of the rules to these variations. The *degree of fulfillment* is a threshold for selecting more effective rules in real-life situations. Figure 8.20 shows an example of fuzzy reservoir operating rules. Details of neural networks and fuzzy set theory are discussed in Chapter 3.

Lund and Guzman (1999) classified the conceptual operating rules for reservoirs in series and in parallel into the categories shown in Tables 8.9 and 8.10. As seen in these tables, some of the conceptual rules for multiple reservoir systems show compatibility between different reservoir purposes, such as flood control, water supply, and energy storage for refill periods on reservoirs in series. However, in many complex systems, it is necessary to use optimization models and methods such as regression, neural network, and fuzzy set theory for defining optimal operating policies.

8.10 COST FUNCTIONS

Reservoir operation cost functions can be classified into two categories:

- Relative cost functions
- Economic based cost functions

TABLE 8.9
Conceptual Rules for Reservoirs in Series

Purpose	Refill Period	Drawdown Period
Water supply	Fill upper reservoirs first.	Empty lower reservoirs first.
Flood control	Fill upper reservoirs first.	Empty lower reservoirs first.
Energy storage	Fill upper reservoirs first.	Empty lower reservoirs first.
Hydropower production	Maximize storage in reservoirs with greatest energy production.	Maximize storage in reservoirs with greatest energy production.
Recreation	Not applicable.	Equalize marginal recreation improvement of additional storage among reservoirs.

Source: Lund, J. R. and Guzman, J., *ASCE J. Water Resources Plan. Manage.*, 125(3), 143–-153, 1999. With permission.

TABLE 8.10
Conceptual Rules for Reservoirs in Parallel

Purpose	Refill Period	Drawdown Period
Water supply	Equalize probability of seasonal spill among reservoirs.	Equalize probability of emptying among reservoirs.
Flood control	Leave more storage space in reservoirs subject to flooding.	Not applicable.
Energy storage	Equalize expected value of seasonal energy spill among reservoirs.	For last time step, equalize expected value of refill season energy spill among reservoirs.
Water quality	Equalize expected value of marginal seasonal water quality spill among reservoirs.	For last time step, equalize expected value of refill season water quality spill among reservoirs
Hydropower production	Maximize storage in reservoirs with greatest energy production.	Maximize storage in reservoirs with greatest energy production.
Recreation	Equalize marginal recreation improvement of additional storage among reservoirs.	Equalize marginal recreation improvement of additional storage among reservoirs.

Source: Lund, J. R. and Guzman, J., *ASCE J. Water Resources Plan. Manage.*, 125(3), 143–153, 1999. With permission.

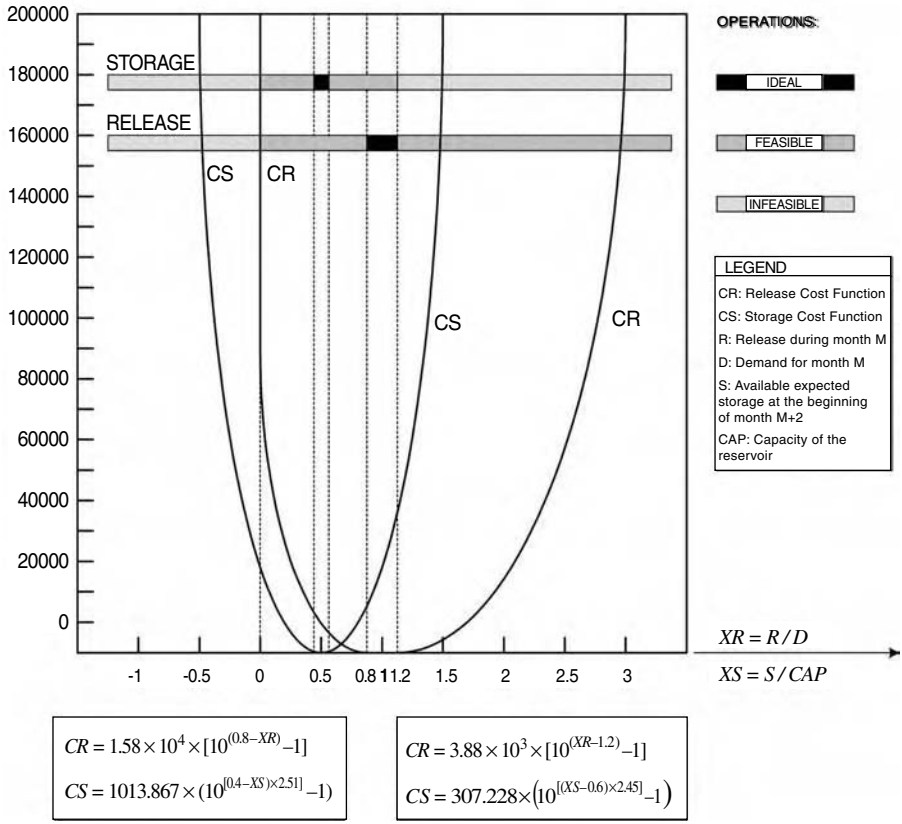


FIGURE 8.21 A typical relative cost function. (From Karamouz, M. and Vasiliadas, H., *Water Resources Research*, 28(5), 1221–1232, 1992. With permission.)

In the relative case, the structure of the cost function reflects the cost of deviation from targets. These targets can be in terms of release, storage, or both. Figure 8.21 shows an example of the ratio of release to demand and storage cost functions. As shown in this figure, the cost functions are convex exponential curves and are developed for comparison purposes. In both sets of cost functions, the x-axis is dimensionless (either release over demand or expected storage over capacity). The cost functions include a specific release safe range (e.g., 80 to 120% of the demand), and a specific storage safe level (e.g., 60 to 80% reservoir capacity for conservation and a higher storage level and perhaps 70 to 85% for generating hydropower). No loss is associated with being in the safe range because:

- Sufficient water is released to satisfy all consumptive needs.
- No flooding occurs because the release is kept at acceptable rates.
- Reservoir storage is kept at a relatively high pool level for conservation and hydropower generation.

Outside the safe range, the losses increase exponentially as the deviations from targets increase.

In the real cost function, the prices of water and power are considered. In estimating the price of power, the following steps should be taken (Karamouz et al., 2002):

- The present value of initial investment is estimated based on the age of the dam and its related facilities, the cumulative interest rate, and rate of depreciation.
- The present value is then distributed over the expected remaining life of the dam on a monthly basis.
- The cost of maintenance and operation should be added to the monthly cost.

The price of power can also be taken into account when considering the duration of power generation in peak and off-peak hours. Attempts to apply actual cost functions usually encounter data insufficiencies, but this type of cost function gives a better picture of the capabilities of operating tools for managers and decision makers in terms of monetary benefits and costs.

8.11 EFFICIENCY OF OPERATING POLICIES

Evaluating the performance of a measuring system is the final step in the application of simulation and optimization models for river–reservoir systems planning and management. The operational status of water resources systems can be classified as satisfactory, unsatisfactory, or failed. Failure can be structural or operational; structural failure of a dam can occur because of flood, earthquake, etc. In this section, operational failure of the system is of concern and we want to evaluate how well the objectives have been satisfied. In other words, the system status in time period t expressed by the random variable X_t can be in set S, which is the set of all satisfactory outputs, or in set F, which is the set of all unsatisfactory outputs.

First, the performance criteria or indices that help the planner to classify the status of a system as S or F should be defined. Even though appropriate definitions for the performance indices depend on the problem and objectives of planning, some basic concepts are similar.

Hashimoto et al. (1982) described systems performance from three different viewpoints:

- How often the system fails (*reliability*)
- How quickly the system returns to a satisfactory state once a failure has occurred (*resiliency*)
- How significant the likely consequences of failure may be (*vulnerability*)

In the following sections, some of the basic performance measures in each of the above categories are presented in order to assess the performance of operation policies for a river–reservoir system.

8.11.1 Reliability

Reliability is the probability that no failure occurs within a fixed period of time:

$$\alpha = \text{Prob}\left[X_t \in S\right] \qquad \forall t \tag{8.82}$$

Based on this definition, reliability is the opposite of risk, in which the probability of system failure is expressed. Another definition of reliability is the probability that no failure will occur within the planning horizon. This indicator is a measure of system performance in meeting target values, which is an important indicator for analyzing the performance of water resources systems in normal conditions.

In the case of a multipurpose river–reservoir system, reliability can be defined in different ways. For example, if the reservoir supplies water and or energy demands, reliability can be defined by the probability that a specific percentage of water and energy demands is supplied within the planning horizon.

Example 8.9

The monthly reservoir releases in 3 years are shown in Table 8.11. The monthly water demand downstream of the reservoir is 10 million m^3. Estimate the reliability of supplying 100% and >80% of demands.

Solution: The reliability of supplying 100% of the monthly demand (α_{100}) can be estimated as the fraction of time that system performance is satisfactory:

$$\alpha_{100} = \frac{\text{No. of months that 100\% of demands are supplied}}{\text{Total No. of Months}} = \frac{19}{36} = 52.7\%$$

The reliability of supplying at least 80% of the monthly demands can also be estimated as follows:

$$\alpha_{80} = \frac{\text{No. of months that more than 80\% of demands are supplied}}{\text{Total No. of Months}} = \frac{28}{36} = 77.7\%$$

It should be noted that the above approach can be used only when the time period is long enough that the frequency of occurrence of a sample approaches the probability.

8.11.2 Resiliency

Resiliency describes how quickly a system recovers from failure once failure has occurred. The resiliency of a system can be considered in the planning horizon as follows:

$$\beta = \text{Prob}\left\{X_{t+1} \in S \middle| X_t \in F\right\} \tag{8.83}$$

TABLE 8.11
Reservoir Releases for a 3-Year Period (Example 8.9)

Month	Year 1 (million m^3)	Year 2 (million m^3)	Year 3 (million m^3)
January	10	8	9
February	12	10	11
March	11	11	10
April	8	8	8
May	7	6	5
June	7	7	6
July	8	9	5
August	10	10	7
September	11	11	9
October	10	9	10
November	11	10	12
December	13	12	11
Annual deficit	10	13	21

As can be seen, resiliency is basically a measure of the duration of an unsatisfactory condition. This indicator is important in drought and flood conditions because damages and costs associated with floods and droughts are affected greatly by the duration of unsatisfactory operations.

Example 8.10

Estimate the resiliency for the river–reservoir system of Example 8.9.

Solution: As shown in Table 8.11, the transition from unsatisfactory to satisfactory states has occurred in the following months:

Year 1: July
Year 2: January, July, and October
Year 3: January and September

Therefore, the probability shown in Eq. (8.83) can be estimated as:

$$\beta = \frac{6}{17} = 35\%$$

If the duration of unsatisfactory condition is needed, then the inverse of β is the average duration of the unsatisfactory condition.

8.11.3 Vulnerability

Vulnerability measures the possible magnitude of a failure if one occurs. For measuring vulnerability, a severity index should be defined. For example, when the main

objective of a river–reservoir system is satisfying water demands, the severity index (s_j) can be defined as the volume of shortage in each time interval. Hashimoto et al. (1982) defined overall system vulnerability as follows:

$$\upsilon = \sum_{j \in F} s_j \, e_j \tag{8.84}$$

where e_j is the probability that x_j, corresponding to s_j, is the most unsatisfactory and severe outcome that occurs among a set of unsatisfactory states. Another definition for vulnerability was presented by Datta and Burges (1984), who defined the vulnerability of a system as the total deficit in the planning time horizon.

Example 8.11

Estimate the vulnerability for the river–reservoir system of Example 8.9.

Solution: The vulnerability can be estimated based on the total deficit occurring within the planning horizon. As shown in Table 8.11, the total deficit of the system can be estimated as:

$$\upsilon = 10 + 13 + 21 = 44 \text{ million m}^3$$

8.12 CONFLICT ISSUES IN RIVER–RESERVOIR SYSTEMS

Reservoirs usually serve multiple objectives, many of which may be in conflict at any given time. For example:

- Several consumptive demand points are located downstream of the reservoirs, and water supplied to one of these points cannot be used by others; therefore, the major conflict issue in reservoir operation in this case occurs when the reservoir is not capable of supplying all of the demands.
- Keeping enough flood control storage conflicts with water conservation strategies for storing water in high-flow seasons.
- Supplying instream requirements conflicts with water conservation and supply objectives.
- Monthly variations of power loads (energy demand) and water demands do not follow the same pattern; therefore, supplying one of these demands in a specific period of time might conflict with supplying another in the future.

In the first item, the main conflict issue in river–reservoir systems planning and operation happens when the reservoir should supply water to various demand points for different purposes. In order to formulate this problem, consider a river–reservoir system that supplies the following demands:

- Domestic water demand
- Agricultural water demand
- Industrial water demand
- Instream flows and environmental water rights for health care

Even though water resources planning for reservoir operation is done by a specific agency (called here the Department of Water Supply) on a regional basis, the water allocation schemes defined and imposed by this department are highly affected by other agencies, such as:

- Department of Agriculture
- Department of Industries
- Department of Environmental Protection

The priorities and favorable ranges of water supply for each of these agencies should be considered in formulating the conflict resolution problems; therefore, the first step in conflict resolution studies is to recognize all of the conflict issues and the responsible agencies.

Each of these departments has its own set of priorities for allocating water to different demands. In the second step, the relative priority of supplying water demands for all related agencies should be defined. For example, irrigation demands have the highest priority for the Department of Agriculture. Other demands, such as industrial water demand, might be a second priority for this department because of the water required for agro-industrial setups.

The next step is to define the acceptable range of demand based on the priority of demands for each agency. For example, the Department of Agriculture could define the utility function for a range of water allocation for irrigation purposes as shown in Figure 8.22. The most favorable annual volume of water allocation for this organization ranges from V_b to V_c million m^3. This range can be estimated based on the variability of irrigation demands. Variations in irrigation demands are a function of variations of net crop water demand due to climate change and uncertainties associated with market price shifts and other socioeconomic factors, as well as the prospect of future development of the agricultural sector. More details about water demand variations are presented in Chapter 10.

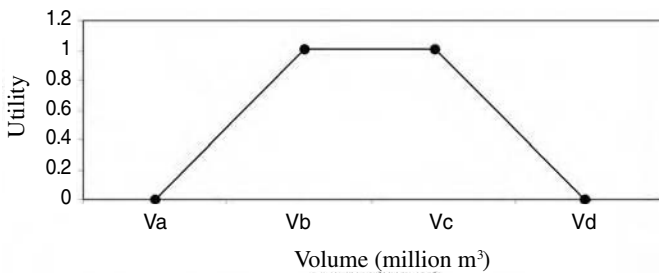


FIGURE 8.22 Utility function of different agencies for water allocated to irrigation demands.

The increasing segment on the left-hand side of this figure shows the range of least favorable to most favorable volume of water that might be allocated to irrigation projects. The points on this line are not totally rejected by this organization because the shortage can be supplied from other sources, such as more expensive groundwater resources. The decreasing segment on the right-hand side of this graph shows that even though the water allocated to irrigation projects supplies all the demands, these conditions are also less favorable because of probable flood damages and violation of water conservation concerns, which might cause shortages in the future.

All the agencies selected in the first step should provide their favorable range of water supply for their various demands. Besides different water demands and priorities of agencies in supplying those demands, each agency has a specific level of authority with respect to changing the water allocation schemes and imposing a favorable range of water supply in the political and institutional climate of each region. For example, the Department of Agriculture has an interest in supplying a relatively higher volume of water to the Department of Industries than the Department of Water Supply. The Department of Industries is comprised of major agencies having higher authority in industrial water allocation schemes. Analysts can also define a set of relative weights based on the authority of these agencies with regard to the final decision on water allocation. Different agencies can also affect the final decision based on the social and political climate of each region.

The final step is to formulate the conflict resolution problem. For this purpose, different methods can be used (as briefly explained in Chapter 2). The steps that should be taken for conflict resolution in river–reservoirs systems modeling can be summarized as follows:

1. The conflict issues should be recognized based on the operation objectives of the river–reservoir systems.
2. The agencies engaged in water resources allocation and consumption should be recognized.
3. The agencies considered in step 2 should provide their own sets of favorable ranges of water supply to different demands.
4. The analyst might define the relative authority of agencies in imposing the allocated water to each demand.

The formulation of a conflict resolution problem in river–reservoir systems is as follows:

$$\text{Maximize} \quad U(x_1, x_2, x_3, x_4) = \sum_{j=1}^{4} \left(\sum_{i=1}^{4} (\mu_{i,j}(x_i) \times W_{i,j}) \right) \tag{8.85}$$

$$\text{Subject to:} \quad x_1 + x_2 + x_3 + x_4 = R \tag{8.86}$$

where:

x_1, x_2, x_3, x_4 are the annual water allocation to domestic, agricultural, industrial, and environmental demands, respectively (million m^3).
$W_{i,j}$ is the assigned weight to agency i in imposing allocated water to demand j.
$\mu_{i,j}(x)$ is the degree of acceptance of agency i if x units of water are allocated to demand j.
$U(X_1, X_2, X_3, X_4)$ is the utility function of the system if x_1, x_2, x_3, and x_4 units of water are supplied to demands 1, 2, 3, and 4, respectively.
R is the annual release from the reservoir (million m^3)

Example 8.12

In a river–reservoir system, the following agencies are affected by the decisions made to release water from the reservoir:

- Agency 1: Department of Water Supply
- Agency 2: Department of Agriculture
- Agency 3: Department of Industries
- Agency 4: Department of Environmental Protection

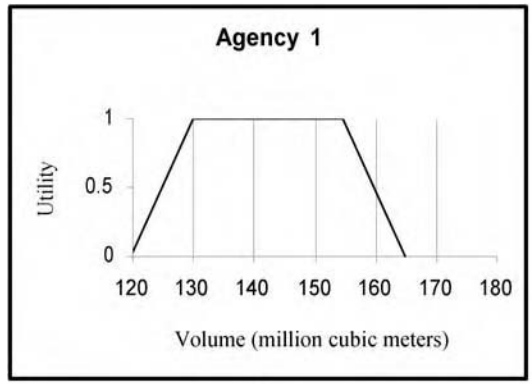
The main objective is to meet the water demand that is of concern to these agencies. The Department of Water Supply has a twofold role — namely, to allocate water to different purposes as well as to supply water for domestic purposes (*DD*). The irrigation demand (*RD*), industrial demand (*ID*), and instream flow (environmental) water demand (*IF*) should also be provided. The decision makers in these agencies are asked to set the most favorable range of water allocation for each demand by assigning a 1 (most favorable) or 0 (least favorable) (see Figures 8.23 to 8.26). The analyst sets the following weights to the role and authority of each agency keeping in mind the political climate of that region:

$$MATA = \begin{array}{c} \text{Agency} \\ DD \\ ID \\ RD \\ IF \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 5 & 1 & 1 & 3 \\ 5 & 2 & 5 & 2 \\ 5 & 5 & 2 & 2 \\ 4 & 2 & 2 & 5 \end{bmatrix} \end{array}$$

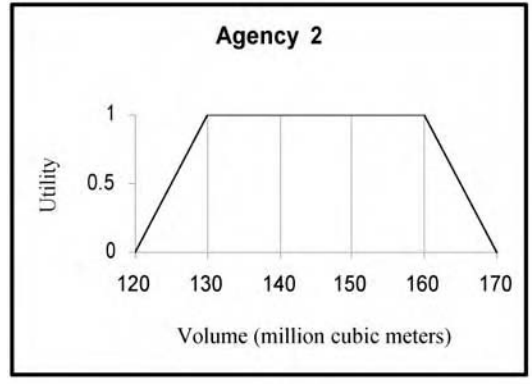
where *MATA* is the matrix of relative authority of agencies.

The average annual inflow to the reservoir is estimated as 1400 million m^3. Find the most appropriate water allocation scheme for a normal year and years with the projected inflows of 1000, 1200, and 1300 million m^3.

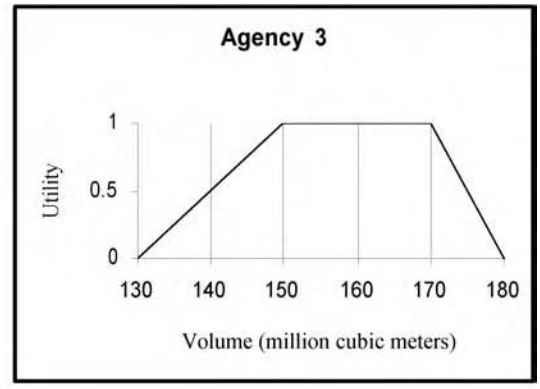
Solution: In order to use Eq. (8.85), the values of $W_{i,j}$ are estimated by normalizing the numbers in the rows of matrix MATA. The normalized values are as follows:



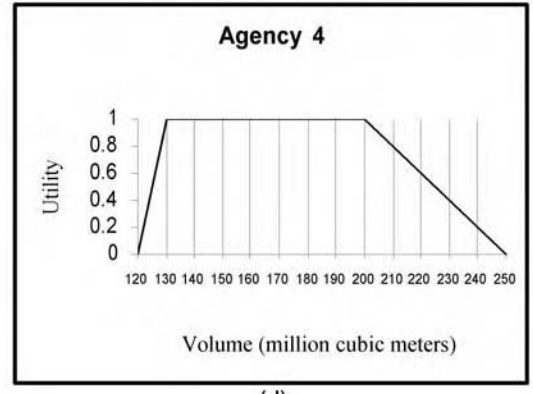
(a)



(b)



(c)



(d)

FIGURE 8.23 Utility function of different agencies for water allocated to domestic demand.

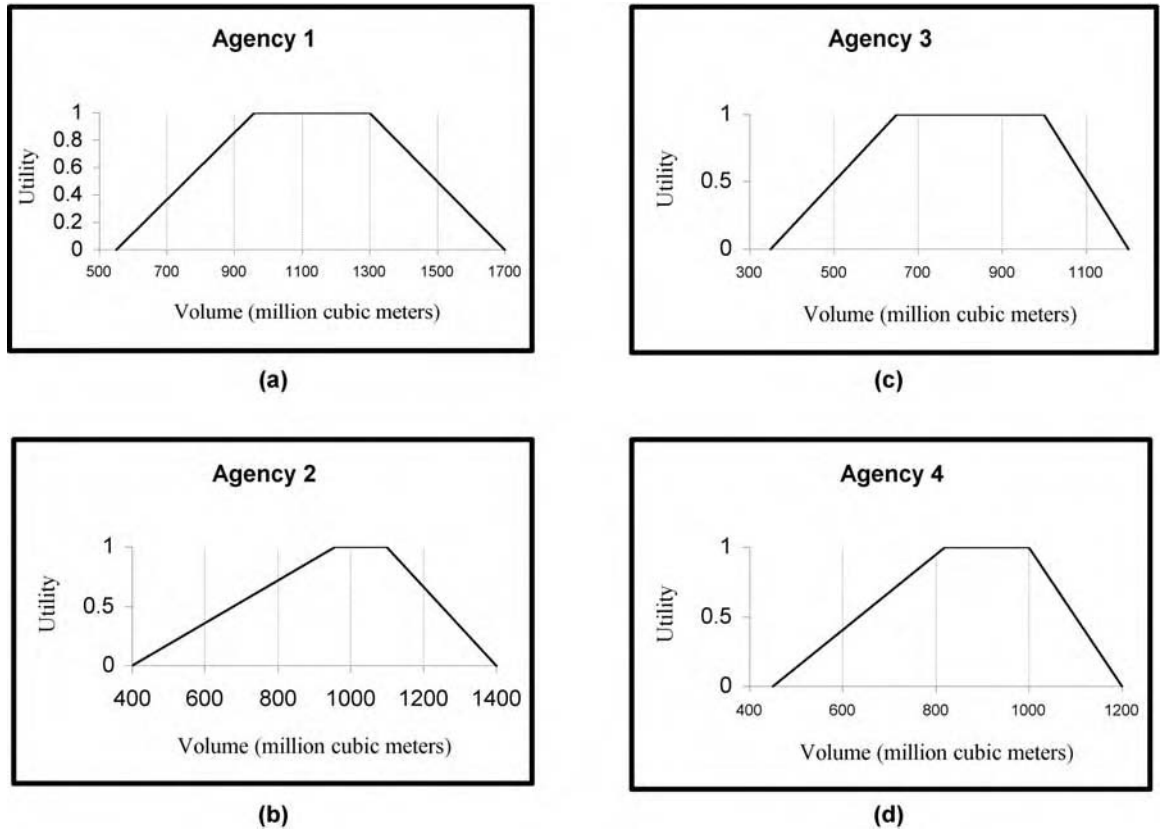


FIGURE 8.24 Utility function of different agencies for water allocated to agricultural demand.

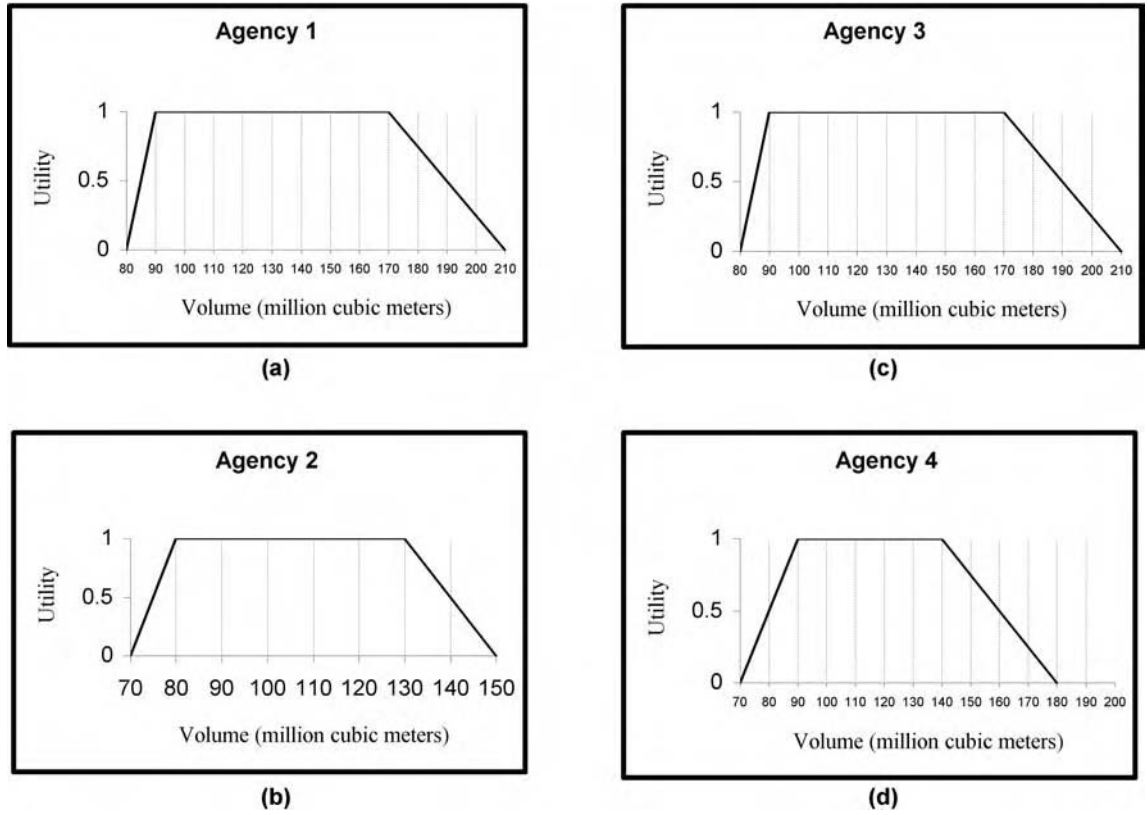
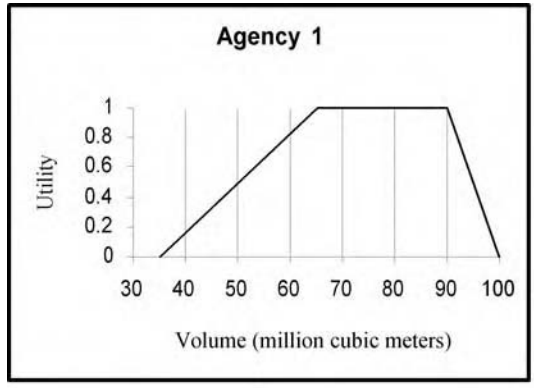
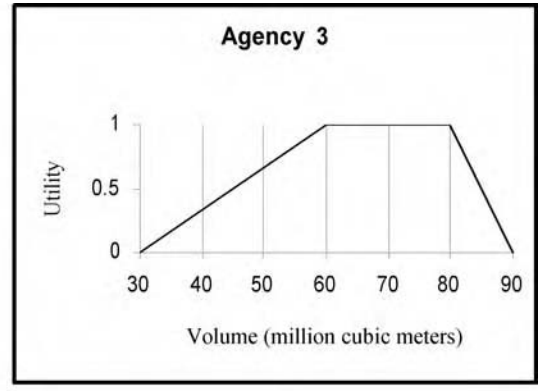


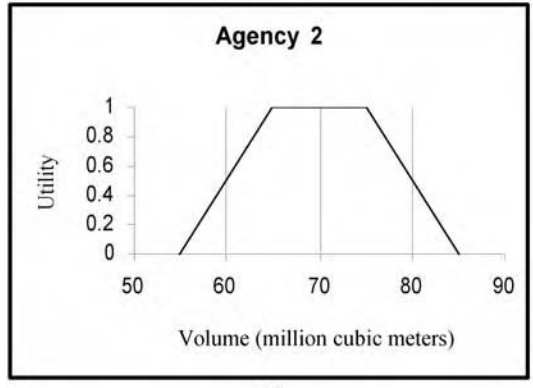
FIGURE 8.25 Utility function of different agencies for water allocated to industrial demand.



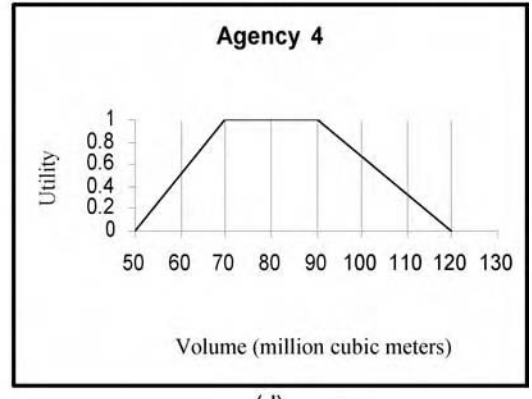
(a)



(c)



(b)



(d)

FIGURE 8.26 Utility function of different agencies for water allocated to environmental water demand.

$$MATA = \begin{array}{c} \\ DD \\ ID \\ RD \\ IF \end{array} \begin{array}{c} \begin{array}{cccc} \text{Agency: } 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0.5 & 0.1 & 0.1 & 0.3 \\ 0.357 & 0.142 & 0.357 & 0.143 \\ 0.357 & 0.357 & 0.143 & 0.143 \\ 0.308 & 0.154 & 0.154 & 0.385 \end{bmatrix} \end{array}$$

$\mu_{i,j}$ can also be computed using Figures 8.23 to 8.26. For example, allocating 1200 million m^3 to irrigation demand is the most favorable option for agency 2 (Department of Agriculture), while it is the least favorable option for agency 4 (Department of Environmental Protection). The linear programming technique is used for solving this problem assuming that release R is equal to inflows of 1000, 1200, 1300, and 1400 million m^3. The results of water allocation are shown in Table 8.12.

For this particular system, the annual water demand for the above purposes has been estimated by an independent party to be, at a minimum, about 150, 960, 110, and 70 million m^3 for the four agencies, respectively, and the agencies are aware of these estimates. As can be seen in the table, in the dry year scenario (inflow = 1000), the domestic and environmental demands have been supplied, while the shortage is imposed on industrial and agricultural demands. By increasing the annual inflow to 1200 and 1300 more water is allocated to the agricultural demands while some shortages are still imposed on industrial sector. When the inflow is increased to 1400 million m^3 all of the demands are supplied.

Another application of MCDM methods in solving conflict resolution problems can be formulated by minimizing the distance from the most favorable point associated with each objective. In case of water supply to different demands from the reservoir, f_i can be estimated based on the authority of each decision maker (responsible agency) as follows:

$$f_i(x_i) = \sum_{j=1}^{m} \left(\mu_{i,j}(x_i) \times W_{i,j} \right) \tag{8.87}$$

TABLE 8.12
Results of Linear Programming for Finding the Most Appropriate Water Allocation Scheme (Example 8.12)

Demand	R = 1000	R = 1200	R = 1300	R = 1400
Domestic	150	150	150	155
Agricultural	690	890	990	1040
Industrial	90	90	90	130
Environmental	70	70	70	75

where:

$W_{i,j}$ is the assigned weight to agency i in imposing allocated water to demand j.
$\mu_{i,j}(x_i)$ is the degree of acceptance of agency i if x_i units of water are allocated to demand j.
m is the total number of decision makers.

The minimum distance from the most favorable points can be found as follows:

$$\text{Minimize} \quad d \tag{8.88}$$

$$\text{Subject to:} \quad w_1 \cdot \left(f_1^*(x_1) - f_1(x_1)\right) \le d \tag{8.89}$$

$$\vdots$$

$$w_n \cdot \left(f_n^*(x_n) - f_n(x_n)\right) \le d \tag{8.90}$$

where:

d is the distance from the most favorable points associated with different objectives.
$f_i^*(x_i)$ is the favorable point for objective i if x_i units of water are allocated to demand i.
$f_i(x_i)$ is the value of objective function i if x_i units of water are allocated to demand i.
w_i is the relative weight of objective i.

By solving the above problem, the optimal values for decision variables $(x_1, \cdots, x_n)$, which are the volume of water allocated to different demands, can be found.

Example 8.13

In the river–reservoir system explained in Example 8.12, find the optimal allocation of 1400 and 1000 million m^3 released from the reservoir using the most favorable point method. In this problem, the four agencies decide on water allocation schemes. Therefore, Eqs. 8.88 to 8.90 can be used as follows:

$$\text{Minimize} \quad d$$

$$\text{Subject to:} \quad \left(f_1^*(x_1) - f_1(x_1)\right) \le d$$

$$\left(f_2^*(x_2) - f_2(x_2)\right) \le d$$

$$\left(f_3^*(x_3) - f_3(x_3)\right) \le d$$

$$\left(f_4^*(x_4) - f_4(x_4)\right) \le d$$

TABLE 8.13
Results of the Most Favorable Point Method for Finding the Most Appropriate Water Allocation Scheme (Example 8.13)

Demand	R = 1400 (million m^3)	R = 1000 (million m^3)
Domestic	176	140
Agricultural	1008	708
Industrial	151	86
Environmental	65	66

Because no specific priority for each demand exists, the w_i values are set to 1. The f_i values are estimated using Eq. (8.87), the utility function for different demands shown in Figures 8.23(a), 8.24(b), 8.25(c), and 8.26(d), and the normalized value of the relative authority of the agencies as shown in the normalized MATA. The results of water allocation are shown in Table 8.13.

As mentioned in the previous example, the annual water demand for the above purposes has been estimated by an independent party to be, at a minimum, about 150, 960, 110, and 70 million m^3, respectively, and the agencies are aware of these estimates. As shown in the table, in the dry year scenario (inflow = 1000), none of the demands has been completely supplied but a more balanced level of shortage is achieved compared with the previous method. Higher levels of shortages (22 to 27%) are applied to industrial and agricultural demands, while the shortage in supplying domestic and environmental demands is limited to 7%.

8.13 PROBLEMS

8.1 Consider reservoir A, which supplies water for domestic, irrigation, and industrial purposes with monthly demands of $D^d(m)$, $D^{ir}(m)$, and $D^i(m)$ million m^3 (m = 1, …, 12), respectively. The carrying capacity of the downstream river is also important in the high flow months. The damage due to reservoir releases (R_t, t = 1, …, N) higher than the carrying capacity of the river (R^T) is estimated as the difference between releases and the river carrying capacity. The benefits of supplying demands are estimated based on the price of water allocated to different demands, which are P^d, P^{ir}, and P^i for domestic, irrigation, and industrial purposes, respectively. Formulate the optimization model for finding the optimal monthly releases in T months when the reservoir inflow is I_t.

8.2 In problem 8.1, if a hydropower plant is constructed to produce electricity from the reservoir releases, formulate the optimization model for finding the optimal release from the reservoir. Consider the benefits of power generation based on the price of power (P^p) and the discharge capacity of the power plant (Q^p).

8.3 In problem 8.1, if the minimum discharge required downstream of the reservoir (after supplying water to demand points) is estimated to be Q_{min} and the minimum reservoir storage is limited to S_{min} for recreation purposes, show how the constraints of the optimization model change.

8.4 The variation of annual water demand of a city over a 30-year planning time horizon is shown in the following table:

	Year				
	2010	**2020**	**2025**	**2035**	**2040**
Water demand (million m^3)	100	120	130	145	170

The current demand of this city is about 90 million m^3. Three projects are studied to supply the demands of this city. Considering the cost of each project and a 3% rate of return, find the optimal sequence of implementing projects using linear programming:

Project	**Capital Investment ($\$ \times 10^6$)**	**Capacity (million m^3)**
Dam A	5	35
Interbasin water transfer tunnel B	8	20
Dam C	7	35

8.5 Suppose you want to design the spillway of a dam. In the following table, the variation between the flood return period and capital investment estimated based on the required capacity of the spillway is shown. The total damages associated with floods during various return periods are also shown in this table.

Return Period (years)	**Annual Probability of Exceeding Capacity**	**Capital Cost (units/year)**	**Damage (units)**
100	0.01	10.05	0
200	0.005	10.6	250
500	0.002	16.05	5000
1000	0.001	22.70	7500
10000	0.0001	30.05	19500

(a) Find the design return period of the spillway using the reliability-based approach.

(b) What is the risk of failure for your design?

8.6 Solve problem 6-4 by using dynamic programming.

8.7 Consider a reservoir that supplies water to an industrial complex. The monthly water demand of this complex is 10 million m^3. The total capacity of the reservoir is 30 million m^3. Let S_t (reservoir storage at the beginning

of month t) take the discrete values of 0, 10, 20, and 30. The cost of operation (*Loss*) can be estimated as a function of the difference between release (R_t) and water demand as follows:

$$Loss_t = \begin{cases} 0 & R_t \geq 10 \\ (10 - R_t)^2 & R_t > 10 \end{cases}$$

(a) Formulate a forward-moving deterministic dynamic programming model for finding the optimal release in the next 3 months.
(b) Formulate a backward-moving deterministic dynamic programming model for finding the optimal release in the next 3 months.
(c) Solve the DP model developed in part (a), assuming that the inflows to the reservoir in the next three months (t = 1, 2, 3) are forecast to be 10, 50, and 20, respectively. The reservoir storage at the current month is 20 million m^3.

8.8 The cumulative density function (CDF) of inflow to a reservoir (Q) is $F_Q(q) = q/(q+1)$. Find the deterministic equivalent of the following constraints:
(a) Prob[$x \leq Q$] ≥ 0.90
(b) Prob[$x \leq Q$] ≤ 0.95
(c) Prob[$x \geq Q$] ≥ 0.75
(d) Prob[$x \geq Q$] ≤ 0.80

8.9 The annual inflow to a reservoir (q) is classified as low ($5 \leq q < 15$), medium ($15 \leq q < 25$), or high ($25 \leq q < 35$). The reservoir inflow transition probabilities are estimated as follows:

$$\text{Prob}\left(q_{t+1} | q_t\right) = \begin{array}{c} \\ \text{low} \\ \text{med.} \\ \text{high} \end{array} \begin{array}{c} \begin{array}{ccc} \text{low} & \text{med.} & \text{high} \end{array} \\ \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.3 & 0.3 & 0.4 \\ 0.1 & 0.5 & 0.4 \end{bmatrix} \end{array}$$

(a) If the inflow in the current year is low ($5 \leq q < 15$), how much is the probability of having high inflow ($25 \leq q < 35$) 2 years from now?
(b) How much is the probability of having a low, medium, and high flow year a few years from now?

8.10 The historical record of monthly inflows to a reservoir has been as follows:

$$q = \{50, 10, 60, 30, 40, 70, 80, 20, 100, 70\}$$

(a) Consider two intervals for the inflow, $q \leq 50$ and $q > 50$, with characteristic values of 30 and 70, respectively. Calculate the inflow transition probability matrix.

(b) Suppose two storage levels for the reservoir are 40 and 60 million m^3. The target volume of monthly release from the reservoir is 20 million m^3. If the cost of operation is estimated to be equal to the difference between the release and the target release (R_t – 20), formulate a backward-moving stochastic dynamic programming model for finding the optimal releases from the reservoir.

(c) Find the stationary policy.

8.11 In problem 8.1, suppose reservoir B is constructed upstream of reservoir A. A stream discharges to the river between reservoirs A and B with monthly discharges of L_t (t = 1, …, N).

(a) Formulate the optimization model for finding optimal releases from reservoirs A and B.

(b) Solve the problem using linear programming and the following assumptions:

Month	$D^d(m)$ (million m^3)	$D^{ir}(m)$ (million m^3)	$D^i(m)$ (million m^3)	Inflow to Dam B (million m^3)	L_t (million m^3)
January	7	10	5	18	5
February	7	20	5.5	20	6.5
March	7	22	5.5	22	7
April	7.3	25	5.5	24	5
May	7.5	15	5	20	4
June	8	10	5	19	4
July	8.2	10	4.5	19	3.5
August	8.1	5	4.5	18	3
September	7.4	0	4.5	17	3
October	7.5	0	4.5	15	3
November	7.1	0	3	15	3.5
December	7	0	3	17	4

R^T = 11.4 cms, P^d = 0.8 \$/m^3, P^{ir} = 0.4 \$/m^3, and P^i = 1.1 \$/m^3

8.12 In a river–reservoir system, the following agencies are affected by the decisions made to release water from the reservoir:

- Agency 1: Department of Water Supply
- Agency 2: Department of Energy

The main objective is to meet the water and energy demands that are of concern to these agencies. The Department of Water Supply has a twofold role — namely, to release water from the reservoir for different purposes as well as to supply water for domestic purposes (*DD*). The main concern for the Department of Energy is to supply power loads (*PL*). A run-of-river power plant downstream of the reservoir, which in turn is located downstream from the diversion point for domestic demands, is operated for power generation. Therefore, that portion of the release from the reservoir that is not supplied to the domestic demand can be used for

power generation. The decision makers in these agencies are asked to set the most favorable range of water and energy supply for each purpose by assigning 1 as most favorable and 0 as least favorable. The following ranges are assigned by the decision makers (see Figure 8.22):

The utility function for the range of water allocation to domestic demand

$$\text{Department of Water Supply}\begin{cases} V_a = 100 \\ V_b = 110 \\ V_c = 140 \\ V_d = 160 \end{cases} \qquad \text{Department of Energy}\begin{cases} V_a = 80 \\ V_b = 100 \\ V_c = 130 \\ V_d = 140 \end{cases}$$

The utility function for the range of water allocation for energy supply

$$\text{Department of Water Supply}\begin{cases} V_a = 200 \\ V_b = 250 \\ V_c = 300 \\ V_d = 320 \end{cases} \qquad \text{Department of Energy}\begin{cases} V_a = 250 \\ V_b = 300 \\ V_c = 350 \\ V_d = 450 \end{cases}$$

The following weights are set for the role and authority of these agencies in allocating water to different purposes:

$$MATA = \begin{array}{c} \text{Agency: } 1 \quad 2 \\ \begin{matrix} DD \\ PL \end{matrix} \begin{bmatrix} 5 & 1 \\ 4 & 5 \end{bmatrix} \end{array}$$

The average annual inflow to the reservoir is estimated to be 400 million m^3. Find the most appropriate water allocation scheme.

REFERENCES

Bellman, R., *Dynamic Programming*, Princeton University Press, Princeton, NJ, 1957.

Bhaskar, N.R. and Whitlatch, E. E., Derivation of monthly reservoir release policies, *Water Resources Res.*, 16(6), 987–993, 1980.

Datta, B. and Burges, S. J., Short-term single multiple purpose reservoir operation: importance of loss function and forecast errors, *Water Resources Res.*, 20(9), 1167–1176, 1984.

Hasebe, M. and Nagayama, Y., Application of fuzzy set theory and neural networks to the control system of reservoir operation from the view point of water resource, in *Stochastic Hydraulic '96: Proc. of the Seventh IAHR Int. Symp.*, Mackay, Queensland, Australia, July 29–31, 1996.

Hashimoto, T., Stedinger, J. R., and Loucks, D. P., Reliability, resiliency, and vulnerability criteria for water resources performance evaluation, *Water Resources Res.*, 18(1), 14–20, 1982.

Hydrologic Engineering Center, *HEC-5 Simulation of Flood Control and Conservation Systems*, U.S. Army Corps of Engineers, Davis, CA, 1982.

Karamouz, M., Houck, M. H., and Delleur, G. W., Optimization and simulation of multiple reservoir systems, *ASCE J. Water Resources Plan. Management*, 118(1), 71–81, 1992.

Karamouz, M. and Vasiliadis, H., A Bayesian stochastic optimization of reservoir operation using uncertain forecast, *J. Water Resources Res.*, 28(5), 1221–1232, 1992.

Karamouz, M. and Zahraie, B., Nonstructural means of flood control for protection of Esfahan metropolitan area in Iran, in *Proc. of Int. Workshop on Non-Structural Flood Control in Urban Areas*, Sao Paulo, Brazil, 1998.

Karamouz, M., Zahraie, B., Torabi, S., and Shahsavari, M., Integrated water resources planning and management for Tehran metropolitan area in Iran, in *Proc. of the 26th ASCE Annual Water Resources Planning and Management Conf.*, Tempe, AZ, June 6–9, 1999.

Karamouz, M., Zahraie, B., Shahsavari, M., Torabi, S., and Araghi-Nejhad, Sh., An integrated approach to water resources development of Tehran region in Iran, *J. Am. Water Resources Assoc.*, 37(5), 1301–1311, 2001.

Karamouz, M., Zahraie, B., and Sorooshian, S., Seasonal streamflow forecasting using snow budget and climate signals: application to Salt River Basin in Arizona, *ASCE J. Hydrologic Eng.*, 2003 (submitted).

Karamouz, M., Zahraie, B., and Araghi-Nejhad, Sh., Operating hydropower reservoirs: a decision support system framework for Dez and Karoon reservoirs in Iran, *ASCE J. Computing Civil Eng.*, 2003 (submitted).

Kottegoda, N. T. and Rosso, R., *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, McGraw-Hill, New York, 1997.

Linsley, R. K., Franzini, J. B., Freyberg, D. L., and Tchobanoglous, G., *Water Resources Engineering*, McGraw-Hill, New York, 1992.

Loucks, D. P., Stedinger, J. R., and Haith, D. A., *Water Resources Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1981.

Lund, J. R. and Guzman, J., Derived operating policies for reservoirs in series or in parallel, *ASCE J. Water Resources Plan. Manage.*, 125(3), 143–153, 1999.

Mays, L. W., *Water Resources Handbook*, McGraw-Hill, New York, 1996.

Mays, L. W. and Tung, Y. K., *Hydrosystems Engineering and Management*, McGraw-Hill, New York, 1992.

Montanari, A., Rosso, R., and Taqqu, M., A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan, *Water Resources Res.*, 36(5), 2000.

Moran, P. A. P., A probability theory of dams and storage systems, *Aust. J. Appl. Sci.*, 5, 116–124, 1954.

Nash, J., Noncooperative games, *Ann. Math.*, 54, 1951.

Raman, H. and Chandramouli, V., Deriving a general operating policy for reservoir using neural networks, *ASCE J. Water Resources Plan. Manage.*, 122(5), 342–347, 1996.

Salas, J. D., Boes, C. D., and Smith, R. A., Estimation of ARIMA models with seasonal parameters, *Water Resources Res.*, 18, 1006–1010, 1982.

Savarenskiy, A. D., A method for streamflow control computation (in Russian), *Gidrotekh, Stroit.*, 2, 24–28, 1940.

Shrestha, B. P., Duckstein, L. E., and Stokhin, Z., Fuzzy rule-based modeling of reservoir operation, *ASCE J. Water Resources Plan. Manage.*, 122(4), 262–269, 1996.

Solley, W. B., Pierce, R. R., and Perlman, H. A., *Estimated Use of Water Resources in the United States in 1990*, U.S. Geological Survey Circular 1081, Washington, D.C., 1993.

Vasiliadis, H. V. and Karamouz, M., Demand-driven operation of reservoirs using uncertainty-based optimal operating policies, *ASCE J. Water Resources Plan. Manage.*, 120(1), 101–115, 1994.

9 Water Quality Management

INTRODUCTION

The quality of surface and groundwater resources can significantly affect water use in many regions, especially in those that are arid and semi-arid. In regions where water pollutants from human activities have seriously degraded water quality, the main issue in water quality management is to control pollution sources. Definition of the control levels depends on the water quality standards defined for the various water uses. The term *water quality management* implies that water should be managed so that no uses at any location will be detrimental to its use at another location. Water quality management is thereby distinguished from *water quantity management*, which is the engineering of water resources systems so that enough water will be provided to all potential users within a region (Krenkel and Novotny, 1980). Section 9.1 discusses the principles of water quality systems analysis, including the major types of pollutants and their resources, water quality criteria, and water quality monitoring. Water quality management in rivers and reservoirs is discussed in Section 9.2, and Section 9.3 addresses groundwater pollution sources and the principles of groundwater quality management.

9.1 PRINCIPLES OF WATER QUALITY SYSTEMS ANALYSIS

9.1.1 Water Pollutants and Their Principal Sources

When the discharge of wastes disturbs the natural ecological balance of a water body, water pollution occurs. The wide range of water pollutants can be classified into major pollutant categories as shown in Table 9.1, which also includes the principal sources of each category. *Point sources* include domestic sewage and industrial wastes because they are collected and discharged into receiving surface and groundwaters from a single point. Sanitary sewage from homes, commercial establishments, and public institutions is referred to as *domestic sewage*. The pollutant sources are *non-point sources* if the pollutants are discharged to water from multiple points. The major non-point sources can be classified as *agricultural return flows* and *urban runoffs*. Reduction or elimination of point sources of pollution can be implemented by proper treatment processes before discharging to receiving waters, but treatment of non-point effluents usually is not economically feasible.

TABLE 9.1
Major Pollutant Categories and Principal Sources of Pollutants

Pollutant Category	Point Sources		Non-Point Sources	
	Domestic Sewage	Industrial Wastes	Agricultural Runoff	Urban Runoff
Oxygen-demanding material	X	X	X	X
Nutrients	X	X	X	X
Pathogens	X	X	X	X
Suspended solids/sediments	X	X	X	X
Salts		X	X	X
Toxic metals		X		X
Toxic organic chemicals		X	X	
Heat		X		

Source: Davis, M. L. and Cornwell, D. A., *Environmental Engineering*, McGraw-Hill, New York, 1991. With permission.

9.1.1.1 Oxygen-Demanding Material

This category of pollutants consists of materials that can be oxidized, thus consuming dissolved oxygen (DO) during the oxidation process. Human waste and food residues are sources of oxygen-demanding materials in domestic sewage. Depletion of DO can occur in a water body by discharging industrial wastes and agricultural runoffs that consist of organic matters.

9.1.1.2 Nutrients

Nitrates and phosphates, derived from municipal wastewater, are *inorganic nutrients*. These nutrients are necessary for the growth of all living organisms but are classified as pollutants because excessive amounts of them can promote plant and algae growth. Fertilizers and phosphorus-based detergents and domestic wastes are the major sources of nutrients.

9.1.1.3 Pathogens

Bacteria, viruses, and other microorganisms excreted by diseased animals or persons and found in wastewater are *pathogenic organisms*. When pathogens containing sewage are discharged into a water body used for drinking and recreation, they pose a dangerous health hazard for the public.

9.1.1.4 Suspended Solids

The total solids content of a wastewater can be classified as *dissolved* or *suspended solids*. Suspended solids consist of organic or inorganic particles, which are transported by wastewater into a receiving water body. Suspended solids can change the qualitative characteristics of water such as turbidity and oxygen demand.

9.1.1.5 Salts

Whenever salt concentrations exceed the level acceptable for a particular use of water, they are classified as pollutants. Groundwater resources usually have higher levels of salt concentration in comparison to surface water resources because of the accumulation of minerals from soil and rocks.

9.1.1.6 Toxic Metals and Organic Compounds

Many heavy metals and chemicals are toxic to living organisms. Pesticides in agricultural runoff, lead and zinc in urban runoff, and toxic metals and toxic organic substances in industrial wastewater are some examples of toxic pollutant sources.

9.1.1.7 Heat

An increase in water temperature can have a negative impact on the ecosystem. The rate of oxygen depletion increases as the temperature increases. This is important where the oxygen-demanding wastes are discharged to the water bodies. Many industries, such as thermal power plants, are the major sources of heat pollution of water bodies.

9.1.2 Water Quality Criteria and Standards for Water Use

McKee (1960) aptly described the differences between standards and criteria as follows:

> The term *standard* applies to any definite rule, principle, or measure established by authority. The fact that it has been established by an authority makes a standard somewhat rigid, official, or quasi-legal, but this fact does not necessarily mean that the standard is fair, equitable, or based on sound scientific knowledge, for it may have been established somewhat arbitrarily on the basis of inadequate technical data tempered by a caution factor of safety. Where health is involved and where scientific data are sparse, such arbitrary standards may be justified. A *criterion* designates a mean by which anything is tried in forming a correct judgment concerning it. Unlike a standard, it carries no connotation of authority other than of fairness and equity nor does it imply an ideal condition. Where scientific data are being accumulated to serve as yardsticks of water quality, without regard for legal authority, the term *criterion* is most applicable.

The early water quality requirements and standards were defined based on the safety of drinking water. Later, pollution control and health authorities gave more attention to the protection of water resources. The surface water quality standards that are used by water pollution control agencies in various countries are classified as stream standards or effluent standards, or as a combination of both. These standards specify numerical effluent limits for designated uses of water resources. The effluent limits allow discharge of a specific amount of a conventional pollutant and either limit or prohibit emission of toxic pollutants. Although effluent standards are emphasized more than other standards, stream standards are enforced when the capacity of water

to assimilate waste is insufficient for the level of effluent loading. In addition to stream and effluent water quality standards, water quality requirements for various beneficial water uses are also important and are presented in the next section. Water quality requirements for domestic, industrial, and agricultural water supplies and for fish and other aquatic life are also presented in the following discussion.

9.1.2.1 Drinking Water Supply

Quality criteria for drinking water have been presented in many documents. Current guidelines for evaluating the suitability of surface or groundwater resource for public water supply are the regulations mandated by the U.S. Environmental Protection Agency (EPA), Title 40, parts 141 and 143, and by the Safe Drinking Water Act (1974) (Mays, 1996). Based on these guidelines, primary and secondary standards should be established. The *primary standard* is for human health protection and the *secondary standard* implies a regulation that specifies the maximum allowable contamination levels that protects the public welfare and may adversely affect the appearance or odor of water. Table 9.A1 (at the end of the chapter) presents a summary of the EPA's national primary standards.

9.1.2.2 Industrial Water Supply

Because of the large range of industrial processes, the water quality requirements for industrial water supply are industry dependent. The water quality requirements can even be different for various parts of a single industry. Water quality characteristics that exceed those given in Table 9.A2 (at the end of the chapter) would probably not be acceptable to industry. For a detailed description of water quality requirements for industries, the reader is referred to Corbitt (1990).

9.1.2.3 Agricultural Water Supply

Water quality is critical to most agricultural crops. Plant growth and soil characteristics such as permeability can be affected by water impurities. Dispersion of clay soils reduces the size of soil pores and decreases the soil permeability. In most water resources used for irrigation, Ca^{++}, Mg^{++}, and Na^{+} significantly affect the soil dispersion. A widely used relationship for estimation of permeability is the *sodium adsorption ratio* (*SAR*), which can be estimated using the following equation:

$$SAR = \frac{(Na^{+})}{\sqrt{\frac{(Ca^{+2}) + (Mg^{+2})}{2}}} \tag{9.1}$$

where Na^{+}, Ca^{++}, and Mg^{++} are concentrations of sodium, calcium, and magnesium (in eq/m^3), respectively. Acceptable *SAR* values for irrigation water are related to soil characteristics and electrical conductivity of water. Representative values are reported in Table 9.A3 (at the end of the chapter).

TABLE 9.2
Summary of Chemical Characteristics of Sampled Water

Constituent	Equivalent Mass (g/eq)	Concentration (g/m^3)	Concentration (eq/m^3)
Ca^{++}	20.0	1.5	0.075
Mg^{++}	12.15	1.1	0.09
Na^{++}	23.0	141	6.13

Example 9.1

Table 9.2 shows the chemical characteristics of a sample of water taken from a well. Determine the *SAR* for this sample. If the concentration of total dissolved solids (TDS) of this sample is measured to be 300 g/m^3, can the water be used for irrigation?

Solution: The *SAR* value can be computed using Eq. (9.1) as follows:

$$SAR = \frac{6.13}{\sqrt{\left(\frac{0.075 + 0.09}{2}\right)}} = 21.34$$

This is a high *SAR* value. The TDS value is moderate, but according to Table 9.A3, the water would be generally unacceptable for irrigation.

9.1.2.4 Aquatic Life

The main objective of water quality standards for aquatic life is to preserve essential environmental conditions for their survival, reproduction, and growth. Based on water quality investigations, the DO content is usually the most important water quality variable. The DO concentration required for warmwater biological systems generally should be more than 5 mg/L, but for coldwater biota it should not be below 6 mg/L. It is important to note that the establishment of water quality criteria for aquatic life is quite difficult because of the different effects of various pollutants; site-specific research may be required depending upon the complexity of aquatic systems.

9.1.3 Pollutant Material Balance in Surface Water

The mass balance equation for pollutants is the basis for surface water quality modeling. From Reynolds' transport theorem in fluid mechanics, the pollutant material mass balance equation is derived as follows:

$$\frac{Dc}{Dt} = \frac{\partial c}{\partial t} + u\frac{\partial c}{\partial x} + v\frac{\partial c}{\partial y} + w\frac{\partial c}{\partial z} \tag{9.2}$$

where c is the pollutant concentration; x, y, and z are the Cartesian coordinates; and u, v, and w are the respective components of velocity vector in the x, y, and z directions. Concentration c consists of the *dissolved pollutant concentration* (c_d) and *sorbed pollutant concentration* (c_{sp}); c_d is expressed as the solute mass per unit volume of water, and c_{sp} is defined as the mass of solute per unit mass of the suspended solids or sediments. The concentration of suspended solids maybe defined as m_s, the mass of solids per unit volume of water. c_{sp} and c_d are related as:

$$c_{sp} = K_{sp}.c_d \tag{9.3}$$

where K_{sp} is the *partition coefficient*. As the sorption phenomenon is related to the carbon content of particles, the K_{sp} and K_{sc} are related as:

$$K_{sp} = f_{cp} \cdot K_{sc} \tag{9.4}$$

where f_{cp} is the mass fraction of carbon in the particles and K_{sc} is the partition coefficient of the solute between water and carbon. Therefore, sorption into the solid is due to the carbon fraction that it contains (Sincero and Sincero, 1996).

The partition coefficient is the ratio of pollutant concentration in the solid phase to its concentration in the liquid phase; it reflects the distribution of pollutants between water and solids. The total pollutant concentration can be estimated as follows:

$$c = c_d(1 + f_{cp}.K_{sc}.m_s) \tag{9.5}$$

Pollutants can be classified as:

- Conservative pollutants
- Nonconservative pollutants
- Growing pollutants

Conservative pollutants (e.g., TDS) do not decay, but *nonconservative pollutants* decay in a control mass (e.g., biochemical oxygen demand, or BOD). *Growing pollutants* grow inside the control mass (e.g., algae). If distributed pollutant sources are located along the travel path of the control mass, these sources will diffuse into the voids and can be considered as direct input, G, in unit of mass per unit time per unit volume of space. A common example of G is the sediment oxygen demand (SOD). Using Eqs. (9.2) to (9.5), the complete material balance equation can be derived as follows (Sincero and Sincero, 1996):

$$\frac{\partial c_d}{\partial t} + \frac{\partial}{\partial x}\left(E_{xx}\frac{\partial c_d}{\partial x}\right) + \frac{\partial}{\partial y}\left(E_{yy}\frac{\partial c_d}{\partial y}\right) + \frac{\partial}{\partial z}\left(E_{zz}\frac{\partial c_d}{\partial z}\right) + \frac{1}{1 + f_{cp}.K_{sc}.m_s}\left(u\frac{\partial c_d}{\partial x} + \upsilon\frac{\partial c_d}{\partial z} + \omega\frac{\partial c_d}{\partial z}\right) = 0 \qquad \text{conservative substances} \tag{9.6}$$

$$\frac{\partial c_d}{\partial t} - \frac{\partial}{\partial x}\left(E_{xx}\frac{\partial c_d}{\partial x}\right) - \frac{\partial}{\partial y}\left(E_{yy}\frac{\partial c_d}{\partial y}\right) - \frac{\partial}{\partial z}\left(E_{zz}\frac{\partial c_d}{\partial z}\right) + \frac{1}{1+f_{cp}.K_{sc}.m_s}\left(u\frac{\partial c_d}{\partial x} + \upsilon\frac{\partial c_d}{\partial y} + \omega\frac{\partial c_d}{\partial z}\right) + kc_d - G = 0 \quad \text{nonconservative substances} \tag{9.7}$$

$$\frac{\partial c_d}{\partial t} + \frac{\partial}{\partial x}\left(E_{xx}\frac{\partial c_d}{\partial x}\right) + \frac{\partial}{\partial y}\left(E_{yy}\frac{\partial c_d}{\partial y}\right) + \frac{\partial}{\partial y}\left(E_{zz}\frac{\partial c_d}{\partial z}\right) + \frac{1}{1+f_{cp}.K_{sc}.m_s}\left(u\frac{\partial c_d}{\partial x} + \upsilon\frac{\partial c_d}{\partial y} + \omega\frac{\partial c_d}{\partial z}\right) - \mu c_d = 0 \quad \text{growing substances} \tag{9.8}$$

where μ is the growth coefficient, k is the lumped decay coefficient, and E_{xx}, E_{yy}, and E_{zz} are dispersion coefficient in the x, y, and z directions, respectively. It should be noted that in Eqs. (9.7) and (9.8), dispersion is going out of the control mass; therefore the positive sign has been prefixed.

9.1.4 Water Quality Monitoring

The purpose of water quality monitoring is to obtain quantitative information about the qualitative characteristics of water using statistical sampling. In other words, the purpose of water quality monitoring is to define the physical, chemical, and biological characteristics of water. The growing concern for maintaining the quality of surface and groundwater has put significant demands on optimal design of monitoring networks and extraction of suitable information from collected data. Monitoring is one of the most important steps of water quality management, as decisions to be made are based on available information related to water quality characteristics. Houlihan and Lucia (1999) categorized the purposes of water quality monitoring as follows:

1. Detection monitoring programs are used to detect an impact to surface and groundwater quality.
2. Assessment monitoring programs are used to assess the nature and extent of detected contaminants and to collect data that may be needed for remediation of contaminants.
3. Corrective action monitoring programs are used to assess the impact of remediation or pollution control programs on contaminant concentrations.
4. Performance monitoring programs are used to evaluate the effectiveness of each element of a remediation system in meeting its design criteria.

Sanders et al. (1987) proposed general guidelines for the establishment of a monitoring system (Table 9.3). Identifying the information to be obtained and the statistical methods for converting the data to useful information are considered in the first two steps, and the third step deals with the monitoring network design. Monitoring operation plans are specified in the fourth step, while in the fifth step information related to the original monitoring objectives is developed.

TABLE 9.3
Major Steps in the Design of a Water Quality Monitoring System

Step 1: Evaluate Information Expectations
Water quality concerns
Information goals
Monitoring objectives
Step 2: Establish Statistical Design Criteria
Development of hypotheses
Selection of statistical methods
Step 3: Design Monitoring Network
Where to sample
What to measure
How frequently to sample
Step 4: Develop Operation Plans and Procedures
Sampling procedures
Laboratory analysis procedures
Quality control procedures
Data storage and retrieval
Step 5: Develop Information Reporting Procedures
Types and timing of reports
Reporting formats
Distribution of information
Monitoring program evaluation
Are information expectations being met?

Source: Sanders, T. G. et al., *Design of Networks for Monitoring Water Quality*, Water Resources Publications, Littleton, CO, 1987. With permission.

9.1.4.1 Water Quality Variable Selection

Selection of the water quality variables is one of the most important steps in the design and operation of water quality monitoring networks. The quality of a water body is usually characterized by interrelated sets of physical, chemical, and biological variables. Therefore, the relationships between water quality and quantity variables are often complex. Many variables are required to completely describe the quality of a water body, but it is not economically feasible to measure all of them. Therefore, the variables should be ranked, and a minimum number of water quality variables should be selected for sampling. The variables can be scored and ranked based on multiple-criteria–decision-making methods. The variables should be selected in both time and space. One problem with the economical selection of water variables is that the water quality variables are time–space stochastic processes that cannot be recorded continuously by an instrument at a site. Establishing the relationship

between water quality and quantity variables and analyzing the dependencies between water quality variables are necessary to select representative variables. These correlations can be used to produce information for variables that are not regularly monitored.

Sanders et al. (1987) proposed a hierarchical ranking of water quantity and quality variables that can be efficiently used in variable selection for a monitoring network (Figure 9A.1 at the end of the chapter). In this hierarchical classification of variables, the optimal cutoff point for selecting suitable variables is determined by considering the monitoring objectives, sample collection laboratory analysis costs, and the correlations among variables.

9.1.4.2 Location of Water Quality Monitoring Stations

The location of a permanent monitoring station is very important; if the collected samples are not representative of the water body, other activities in the water quality monitoring program can be of no consequence. The factors that can affect the location of a water quality monitoring station are different for groundwater and surface water resources; length is the dominant space coordinate in rivers, while depth and width have the same role in groundwater. Therefore, in the following discussion, the principles of river and groundwater sampling location are discussed separately.

9.1.4.2.1 Location of River Water Quality Sampling Station

The location of river quality sampling stations is related to the monitoring objectives, number of samples to be collected at each station, and monitoring budget. When the monitoring network has more than one objective, samples and stations should be classified based on the different objectives. When important areas (for example, those with a large population) require more extensive water quality monitoring, the monitoring resources can be allocated according to the spatial variation of land use. Sanders et al. (1987), defined three levels of design criteria for river sampling station location:

1. *Macro-location* — selection of river reaches that will be sampled
2. *Micro-location* — selection of a station location within a selected river reach
3. *Representative location* — selection of points in the river cross-section that provide the best lateral water quality profile of the stream

Selection of water quality sampling reach (macro-location). When an entire river or river basin is monitored, some suitable reaches in the river should first be selected for sampling based on the specific objectives of the monitoring. In the classical methods, macro-location of sampling stations is based on the percentage of basin area that belongs to each reach and on land-use variations such as distributions of industries, population, and agricultural areas. Sharp (1971) proposed a systematic procedure for locating sampling stations in a river system based on the number of tributaries. In this method, which identifies and isolates sources of pollution, the river

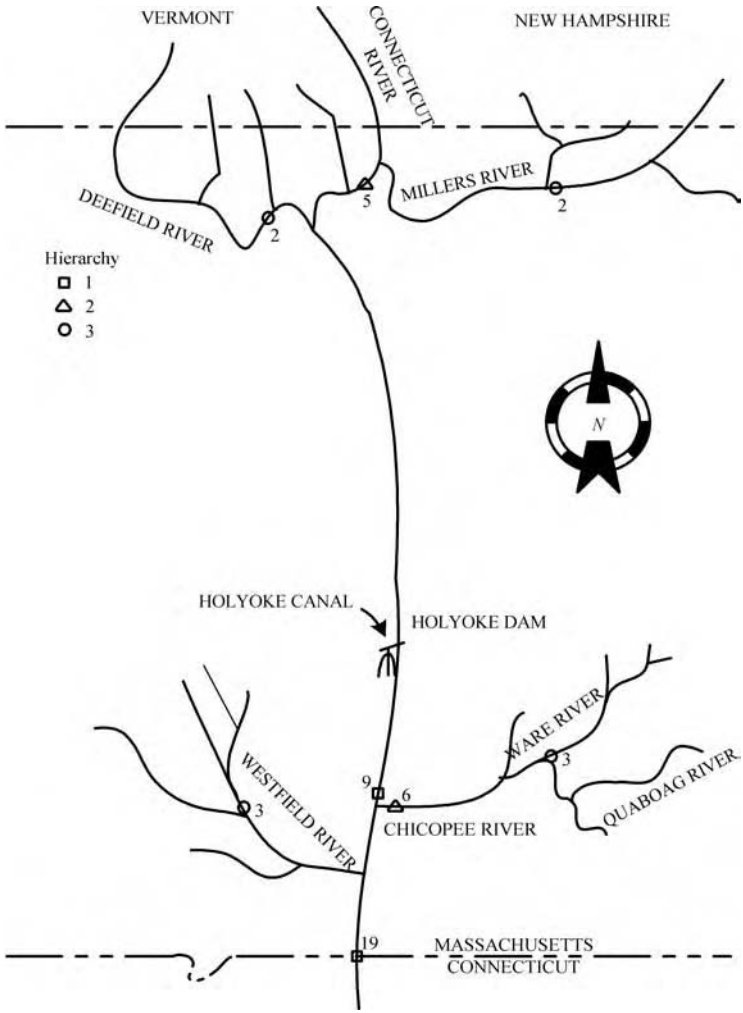


FIGURE 9.1 Location of sampling stations using Sharp's procedure based on the number of tributaries. (From Sanders, T. G. et al., *Design of Networks for Monitoring Water Quality*, Water Resources Publication, Littleton, CO, 1987. With permission.)

system is subdivided into reaches that have relatively equal contributing tributaries. A magnitude (number) of 1 is assigned to the exterior tributaries, and it is assumed that an exterior tributary has a minimum mean discharge. As shown in Figure 9.1, the magnitude of each section of river is equal to the sum of the magnitudes of the intersecting streams. Therefore, the magnitude of the final stretch of the river will be equal to the total number of exterior tributaries. By dividing the magnitude of the final stretch of the river by 2, the centroid of the basin can be determined. The first centroid, which is referred to as the *first-hierarchy sampling reach*, provides two approximately equal portions. Again by dividing by 2, a new centroid can be found

for each part and *second-hierarchy sampling reaches* are defined. More levels of hierarchy are defined by successive subdivisions. As shown in Figure 9.1, the link with the closest magnitude to the centered magnitude is specified as the centroid. Once a link is chosen, a sampling location is selected at its downstream junction. For example, in Figure 9.1, which has three hierarchies, sampling may be done at eight stations: two samples from the first hierarchy, two samples from the second hierarchy, and four samples from the third hierarchy. To identify and isolate a pollution source, Sharp (1971) proposed that samples should be drawn at one hierarchy and analyzed to select the next section of the river that should be sampled (at a sampling station of the next hierarchy), and so on, until all the pollution sources are identified. The magnitudes of tributaries and the main stem can be assigned based on the number of outfalls that discharge into them or by their pollutant load. When two or more variables are measured, more than one water quality monitoring network might be established, and it can be costly and time consuming.

Sampling Station Location within a River Reach (Micro-Location). After the macro-location process and selection of the sampling reaches, sampling sites within each reach should be identified to provide representative samples. This kind of longitudinal sampling location is referred to as micro-location. Micro-location should take place in a reach of river in which the water is completely mixed — that is, where the concentration of water quality variables is independent of lateral location and depth in a cross-section. The distance from an outfall to a mixing zone is a function of the physical characteristics of the river, average stream flow velocity, lateral and vertical dispersion coefficients, and exact location of pollutant source (outfall) in the lateral transect of the river. Sanders et al. (1987) defined mixed distances from a point source for both complete lateral and vertical mixing as follows:

$$L_y = \frac{\sigma^2{}_y}{0.46d} \cdot \frac{u}{u^*}$$
$$L_z = \frac{7.5\sigma^2{}_z}{d} \cdot \frac{u}{u^*} \tag{9.9}$$

$$u^* = \sqrt{gRS_e} \tag{9.10}$$

where:

- L_y is the mixing distance between the point source and complete lateral mixing.
- L_z is the mixing distance between the point source and complete vertical mixing.
- σ_y is the distance between the point of injection and the farthest lateral boundary of river.
- σ_z is the distance between the point of injection and the farthest vertical boundary of river.

d is the depth of flow.
u is the mean stream velocity.
u^* is the shear velocity.
g is the gravity acceleration.
R is the hydraulic radius.
S_e is the slope of the energy gradient.

The calculated mixing distances are valid only for the considered set of hydrologic and hydraulic variables such as stream flow, hydraulic radius, and slope; therefore, the uncertainty of the variables and the estimated mixing distances should be considered and a mixing distance with an acceptable reliability should be determined. Sometimes the river reach does not include a complete mixed zone, for example, when temperature or concentration stratification is present. In this case, determining the sampling location is difficult and water quality concentration variation within a river cross-section should be analyzed using statistical methods such as analysis of variance. When the water quality variation, which can be time dependent, is insignificant, the section is considered to be completely mixed; otherwise, more than one sampling point is required within the cross-section. For a detailed description of the complete mixing assessment of a river cross-section, the reader is referred to Sanders et al. (1987).

Example 9.2

Compute the minimum distance between a pollutant discharge point in a river and the complete mixing zone. The outfall point is located in mid-depth and mid-width in the river cross-section. Assume that the average stream velocity is 0.9 m/sec, the average width is 100 m, and the average depth is 4 m. The slopes of the stream bed and the energy gradient are assumed to be the same and equal to 0.002.

Solution: u^* can be computed using Eq. (9.10):

$$u^* = \sqrt{gRS_e} = \sqrt{9.8 \times 4 \times 0.002} = 0.28 \frac{\text{m}}{\text{s}}$$

Using Eq. (9.9) and assuming that $R \approx 4$ m, it can be written that:

$$L_y = \frac{50^2}{0.46 \times 4} \times \frac{0.9}{0.28} = 4367.23\,\text{m}$$

$$L_z = \frac{7.5 \times 2^2}{4} \times \frac{0.9}{0.28} = 241\,\text{m}$$

As the L_z is less than L_y, the mixing distance from the discharge point is equal to 4367.2 m.

9.1.4.2.2 Groundwater Sampling Location

Groundwater monitoring objectives should be defined before monitoring because they can specify what procedures, techniques, and analyses are required. The main objectives of groundwater monitoring are:

1. Ambient trend monitoring
2. Source monitoring

Ambient trend monitoring is used to discover the spatial and temporal trends of groundwater quality variables. The cost of ambient trend monitoring is very much related to the number of sampling stations and the sampling frequency.

Source monitoring determines the qualitative characteristics and the rate of pollutant movement. The flow rate can be determined using flow net. The mass flow rate of the pollutant can be determined as:

$$R_{total} = \sum_{i=1}^{n} q_i.c_i \tag{9.11}$$

where:

R_{total} is the total mass flow rate.
n is the total number of sections.
q_i is the flow rate in section i.
c_i is the water quality variable concentration in section i.

A suitable groundwater monitoring network should provide the spatial and temporal characteristics of groundwater pollution for various sources of pollution such as plumes.

Hudak and Loaiciga (1993) proposed a quantitative, analytically based hydrologic approach for monitoring network design (sampling location) in a multilayered groundwater flow system. In this method (described below), the susceptibility to contamination of points in the model domain is quantified using weight values. This network design methodology consists of three steps:

1. Domain definition and discretization
2. Determination of relative weights for candidate monitoring sites
3. Selection of an optimal configuration for monitoring network

Definition and Discretization of the Model Domain. The model domain includes the contaminant source and surrounding areas that could be affected by the pollutant source and within which monitoring wells are selected. Zones that encompass the pathways of contaminant migration are defined by advection envelopes. Advection envelopes are extended from hydrogeologic outlets of pollutant sources based upon the known or inferred variation of the hydraulic head (Figure 9.2). The

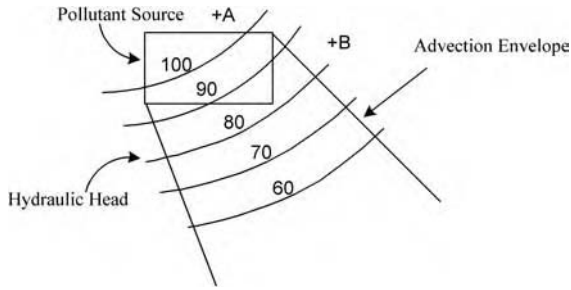


FIGURE 9.2 Advection envelope extended from hypothetical contaminant source. (From Hudak, P. F. and Loaiciga, H. A., *Water Resources Res.*, 29(8), 2835–2845, 1993. With permission.)

two flow lines that originate from separate ends of the hydrogeological outlet can be considered as boundaries of advection envelopes. Monitoring well sites can be classified as *background* and *detection monitoring sites*. Detection monitoring wells are constructed for early detection of contamination by pollutant sources; background monitoring sites determine the background water quality and are located in the upgradient zone of the model domain. The background monitoring wells should not be located adjacent to the contaminant source and advection envelope. They also should not intersect an outward normal line extended from an advection envelope (Hudak and Loaiciga, 1993).

Calculation of Weights for Candidate Monitoring Sites. The relative weight of each monitoring site should be calculated based on the location of the site relative to the pollutant source and advection envelope. The following model can be used for designing a monitoring network for an uncontaminated aquifer (before the pollutant is released from the pollutant source). Therefore, the existing contamination concentrations are not used in the monitoring well configuration. The relative weight of each candidate monitoring well in a multilayer aquifer is determined as (Hudak and Loaiciga, 1993):

$$W_{jk} = \frac{1}{D(s)_{jk} D(e)_{jk}} \tag{9.12}$$

where:

j is a real index of the potential well site
k is the hydrostratigraphic interval (*HSI*) index (*HSI* is defined as a layer that has relatively uniform hydraulic conductivity).
$D(s)_{jk}$ is the horizontal distance from a contaminant source boundary to node j in *HSI* k.
$D(e)_{jk}$ is the perpendicular distance from node j in *HSI* k to the closest boundary of the advection envelope boundary (equal to $D(e)$(max) if node j cannot be intersected by a perpendicular line from an advection envelope).
$D(e)_{jk}$(max) is the maximum value of $D(e)_{jk}$ among nodes for which the quantity can be measured.

Based on the above equation, nodes located near an area with higher probability of contamination have higher weights. To reduce the computational efforts, the boundaries of pollutant source and advection envelopes are approximated by nodes within a grid of candidate groundwater monitoring wells. The value of $D(e)_{jk}$ is calculated only for nodes that are susceptible to contamination from plume spreading and are intersected by an outward normal line extended from an advection envelope. Other nodes that are not intersected by an outward perpendicular line from an advection envelope are assigned the maximum value of $D(e)_{jk}$ calculated for all other nodes. Nodes located within the advection envelope are assigned the minimum value of $D(e)_{jk}$ to avoid division by zero. For example, in Figure 9.2, nodes A and B are located an equal distance from the advection envelope, but node B has the greater probability of being contaminated because node A is located in an upgradient area and should be assigned the maximum of $D(e)_{jk}$ ($D(e)$(max)).

Optimal Configuration of Monitoring Well Network. Optimal configuration of monitoring sites can be obtained using a mathematical model. The objective function of this model is the preferential location of monitoring wells at points with high probabilities of contamination. The model formulation can be as follows (Hudak and Loaiciga, 1993):

$$\text{Max } Z = \sum_{k \in K} \sum_{j \in J_{k-uk}} W_{jk} x_{jk} - \sum_{k \in K} \sum_{j \in J_{uk}} W_{jk} x_{jk} \tag{9.13}$$

$$\text{Subject to } \sum_{j \in J_k} x_{jk} \geq P_k(\min) \quad \text{for each } k \in K \tag{9.14}$$

$$\sum_{j \in J_{uk}} x_{jk} = P_{uk} \qquad \text{for each } k \in K \tag{9.15}$$

$$\sum_{k \in K} \sum_{j \in J_k} x_{jk} = P \tag{9.16}$$

$$x_{jk} = \begin{cases} 0 & \text{otherwise} \\ 1 & \text{if a well is installed at node } j \text{ in } HSI\ k \end{cases} \qquad \text{for each } j \in J_k, k \in K \tag{9.17}$$

where:

j is the areal index of a potential well site.
J_k is the set of potential well sites in *HSI k*.
J_{k-uk} is the set of potential well sites (excluding sites in the upgradient zone) in *HSI k*.
J_{uk} is the set of potential well sites in the upgradient zone in *HSI k*.
k is the *HSI* index.

K is the set of *HSI*s.
W_{jk} is the weight for node j in *HSI* k.
P_k(min) is the minimum number of wells to be located in *HSI* k.
P_{uk} is the number of wells allocated to the upgradient zone in *HSI* k.
P is the total number of wells.

The second term of Eq. (9.13) ensures that the nodes with the lowest weight will be selected for background monitoring in the upgradient zone. Equations (9.14) and (9.15) constrain the minimum and total number of wells in the upgradient zones in each *HSI*. Equation (9.16) ensures that all the wells are located throughout the model domain, and Eq. (9.17) shows that variable x_{jk} is a binary integer.

The total number of monitoring wells in each layer can be determined by budgetary constraints or regulatory requirements. The budget constraint can be written as:

$$\sum_{k\varepsilon K}\sum_{j\varepsilon J_k} C_{jk}x_{jk} \le R \tag{9.18}$$

where C_{jk} is the construction cost of a well at monitoring site j in *HSI* k, and R is the total available budget. The proposed formulation can be solved using integer programming techniques. This model can define the location of the P monitoring wells, including P_{uk} upgradient wells. For more details, the reader is referred to Hudak and Loaiciga (1993) and Loaiciga et al. (1992).

9.1.4.2.3 Sampling Frequency

Sampling frequency shows how often water quality samples should be gathered. Detection of stream standards violations and determination of temporal water quality variations are the main criteria for sampling frequency analysis. The selection of sampling frequency requires some background information on the behavior of the random variables that are important in water quality. The analyst may face three situations:

- Water quality data are available for the stream being studied.
- Water quality data are available only from a stream in a region that has characteristics similar to those of the study area.
- No data exist for the stream under consideration or other similar streams.

When no data are available, a two-stage monitoring program should be utilized. In this case, design data collected in the first stage can be used for suitable design of the monitoring network and to estimate the sampling frequency. When the water quality data are available only for similar rivers in the region, data should be generated for the stream that is under study. In this case, similarities between their geological and hydrologic characteristics, distribution of the pollutant sources, and pollutant characteristics are usually considered for water quality data generation.

9.1.4.2.4 Methods of Selecting Sampling Frequencies

Sampling frequency design can vary from a single-station, single water quality variable to a multiple-station, multiple water quality variable. Sanders et al. (1987) proposed statistical methods for each case which are presented next.

Single-Station and Single Water Quality Variables. In this case, the selected sampling frequency for a specified water quality variable at a station should provide the desired confidence interval around the annual mean. If the collected samples can be assumed to be independent, the variance of sample mean is:

$$Var(\bar{x}) = \frac{\sigma^2}{n} \tag{9.19}$$

where σ^2 is the population variance, and n is the number of samples. The confidence interval for the population mean, μ, is as follows:

$$\bar{x} - Z_{\alpha/2}.\left(\frac{\sigma^2}{n}\right)^{1/2} \le \mu \le \bar{x} + Z_{\alpha/2}.\left(\frac{\sigma^2}{n}\right)^{1/2} \tag{9.20}$$

where $Z_{\alpha/2}$ is the standard normal deviation corresponding to the probability of $\alpha/2$. By rearranging terms, the number of samples (n) required to estimate the mean with a known confidence level (α) is:

$$n \ge \left[\frac{Z_{\alpha/2}.\sigma}{\mu - \bar{x}}\right]^2 \tag{9.21}$$

When the sample standard deviation(s) is used instead of σ, the $Z_{\alpha/2}$ is replaced with Student's t statistic, $t_{\alpha/2}$:

$$n \ge \left[\frac{t_{\alpha/2}.s}{\mu - \bar{x}}\right]^2 \tag{9.22}$$

In this case, the computation of n becomes an iterative problem because when the t distribution is used the number of degrees of freedom ($n - 1$) must be known, but in this case it is not known. Therefore, for solving the problem, an initial value for n is estimated, the degrees of freedom are determined, and a new estimation of n is obtained using Eq. (9.22). By repeating this procedure, the values of n converge to a desired value. When the number of samples is large (more than 30), σ can be estimated with s, without any significant error in determination of the sampling frequency.

Example 9.3

Select the sampling frequency for a station that monitors BOD concentration in an important control point on a river. The annual mean and variance of the BOD concentration based on historical data are 6.01 mg/l and 8.1 $(mg/l)^2$, respectively. The desired confidence interval width is assumed to be 3 mg/l with a 95% confidence level.

Solution:

$$\text{The confidence interval width} = 2\times 1.96\times\frac{\sigma}{\sqrt{n}}$$

$$2\times 1.96\times\frac{\sigma}{\sqrt{n}} = 3$$

$$n = \left[\frac{2\times 1.96}{3}\right]^2 \sigma^2 = \left[\frac{2\times 1.96}{3}\right]^2 \times 8.1$$

$$n = 13.8 \approx 14 \text{ samples / year}$$

Single Station and Multiple Water Quality Variables. If the problem of sampling frequency selection includes more than one water quality variable, a compromise program should be used to obtain the most acceptable confidence interval that considers all water quality variables. A simple way to do this is to compute a weighted average of confidence interval widths for several water quality variables. The relative weights of water quality variables can be selected by engineering judgment so that the contributions of all water quality variables are comparable.

Example 9.4

Select the sampling frequency for a station that has four water quality variables. Assume that the 95% confidence interval width about the mean of the water quality variables will be equal to one fourth of the average of these variables. The historical population statistics of the variables are as follows:

Variable	Mean (mg/l)	Variance ([mg/l]²)
BOD	41	286
Total Nitrogen (TN)	10	18.1
Ca^{++}	40	38
Mg^{++}	34	60

Solution: The width of 95% confidence interval should be one fourth the average of the means. Thus:

$$\frac{R_1 + R_2 + R_3 + R_4}{4} = \frac{1}{4}\left[\frac{\mu_1 + \mu_2 + \mu_3 + \mu_4}{4}\right]$$

where R_i is the width of confidence interval in the station i that can be computed using the following equation:

$$R_i = 2(1.96)\frac{\sigma_i}{\sqrt{n}} \qquad i = 1,2,3,4$$

$$\frac{1}{4}(\mu_1 + \mu_2 + \mu_3 + \mu_4) = \frac{2 \times 1.96}{\sqrt{n}}(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4)$$

$$n = \left[4 \times 2 \times 1.96 \frac{(16.91 + 4.25 + 6.16 + 7.75)}{(41 + 10 + 40 + 34)}\right]^2 = 19.35 \approx 20 \text{ samples / year}$$

Multiple Stations and Single Water Quality Variable. Because the designer usually wants to obtain the same information from each sampling station of the monitoring network, this requirement can be a basis for selection of the sampling frequency. When based on historical data, the average concentrations of water quality variables at the stations are approximately equal, the sampling frequencies should be selected so they provide equal confidence intervals about the means for the different stations. In a simple procedure proposed by Ward et al. (1979), the total number of samples (N) is allocated to stations based on their relative weights using the following equation:

$$n_i = w_i.N \tag{9.23}$$

where n_i is the number of samples at station i, w_i is the relative weight of station i, and N is the total number of samples. The relative weights show the relative importance (priority) of stations and can be estimated based on historical data. For example, relative weights based on historical means or historical variances can be computed using:

$$w_i = \frac{\mu_i}{\sum_{i=1}^{N_s} \mu_i} \tag{9.24}$$

$$w_i = \frac{\sigma_i^{\,2}}{\sum_{i=1}^{N_s} \sigma_i^{\,2}} \tag{9.25}$$

where μ_i and $\sigma_i^{\,2}$ are the historical mean and variance of the water quality variable at station i and N_s is the total number of stations in the water quality monitoring network.

Multiple Stations and Multiple Water Quality Variables. The design problem encountered most frequently in practice involves more than one water quality variable as well as more than one sampling station. In this case, a separate sampling

TABLE 9.4
Means and Variances for Three Water Quality Variables (Example 9.5)

	TN		BOD		Total Phosphorus (TP)	
Station	Mean (mg/l)	Variance	Mean (mg/l)	Variance	Mean (mg/l)	Variance
1	2.4	0.50	4.51	3.41	0.21	0.01
2	3.3	0.90	5.01	8.04	0.22	0.01
3	5.4	1.31	3.50	1.52	1.28	0.01
4	3.8	1.20	4.41	3.66	0.23	0.02

frequency should be selected for each water quality variable at each station and a weighted average of them can be considered as each station's sampling frequency. Sanders et al. (1987) suggested that a weighted average variance for the water quality variables can be computed for each station and used in Eq. (9.25). In the case of multiple variables, it is not possible to obtain equal confidence interval widths for water quality variables in all stations if all variables at a given station are sampled at the same frequency.

Example 9.5

The historical means and variances for three water quality variables in four stations have been calculated and are shown in Table 9.4. Select the sampling frequency for each station using the weighting factors based on historical variances. The total number of samples per year is considered to be equal to 40.

Solution: The total number of samples should be allocated to each station for each water quality variable using weighting factors. The relative weight of each station for each variable is as follows:

$$w_i = \frac{\sigma_i^2}{\sum_{i=1}^{4} \sigma_i^2} \qquad i = 1, 2, 3, 4 \tag{9.26}$$

The weighting factors for the water quality variables are based on the variances and are provided in Table 9.5. Table 9.6 shows the average number of samples estimated for each station.

9.2 SURFACE WATER QUALITY MANAGEMENT

9.2.1 River Water Quality Management

The protection of public health in large cities and near industries located along rivers is the main purpose of river pollution control; however, protection of water resources

TABLE 9.5
The Relative Weights of Stations for Example 9.5

Station	TN	BOD	TP
1	0.13	0.21	0.2
2	0.23	0.48	0.2
3	0.33	0.09	0.2
4	0.31	0.22	0.4

TABLE 9.6
Allocation of Samples via Proportional Sampling Based on Variances

Station Number	Number of Samples Allocated			Average (TN, BOD, and TP)
	TN	BOD	TP	
1	5	8	8	7
2	9	19	8	12
3	13	4	8	8
4	13	9	16	13

and ecosystems and preserving the natural environment related to these resources have become more important in recent years. Therefore, the main objective of river water quality management today is to control pollutants discharged to the river so that river water quality at critical sections is not reduced to an unacceptable extent below the natural background level. Important steps in river water quality management include measurement of discharged pollutants, prediction of the impact of a pollutant on water quality using suitable river water quality simulation models, and determination of water quality criteria.

Pollution of rivers can occur directly from sewer outfalls or industrial discharges as point sources and agricultural or urban runoffs as non-point sources or indirectly from air pollution. Impacts of a pollutant on river water quality depend on the pollutant type and load and river characteristics. The following sections provide an overview of the derivation of descriptive models for the transport of pollutants and the formulation of some important theoretical optimization models for water quality management in rivers and streams.

9.2.2 Low Flow Determination

Estimates of low flow quintiles, such as the 7-day, 10-year low flow, are used widely in steady-state water quality modeling and management. Low flow quintiles are usually obtained by statistical modeling of observed data series. For this purpose, several methods have been presented in the literature. In general, the low flow determination can be stated as the estimation of the quantity q_T in the following equation.

$$F_Q(q_T) = \int_0^{q_T} f_Q(q)dq = \frac{1}{T} \tag{9.27}$$

where T is the recurrence interval, $f_Q(q)$ is a model of the probability density function of the random variable Q. In low flow analysis, T usually is considered to be equal to 10 when Q is the annual minimum 7-day average discharge and q_{10} is the 7-day, 10-year low flow. For details, the reader is referred to Durrans (1996).

The low flow can also be computed using the empirical frequency analysis method. In this method, the smallest flow that occurs for 7 consecutive days in each year is computed using the long-term flow record (n years) for the location being modeled. In the next step, the n flows are tabulated in ascending order and a rank m is assigned to them. Then, the cumulative values of occurrence and recurrence intervals are calculated using the following equations:

$$p = \frac{m}{n+1} \tag{9.28}$$

$$T = \frac{1}{p} \tag{9.29}$$

Application of this method is illustrated in the following example.

Example 9.6

The ranked values of 20 years of data for the 7-day minimum average flows of a river are given in Table 9.7. Prepare the low flow frequency curve and determine the 10-year low flow.

Solution: As can be seen in Figure 9.3 and Table 9.7, a value of about 1.65 m^3/sec corresponds to the 7-day, 10-year low flow.

9.2.3 Computer Models for Streamwater Quality Simulation

Computer models for streamwater quality simulation have been used widely for water quality management of complex stream systems. These models enable water resources engineers to predict the effects of numerous pollutant discharges, and they often can be used to preserve water quality by managing human activities as the main pollutant source. Tables 9.A5 and 9.A6 at the end of the chapter present and compare the important attributes of some surface water quality models. The following discussion summarizes the characteristics of two useful water quality simulation models.

TABLE 9.7
Annual 7-Day Minimum Average Flow in Example 9.6

Rank	7-Day Low Flow (m^3/sec)	Probability	Recurrence Interval	Rank	7-Day Low Flow (m^3/sec)	Probability	Recurrence Interval
1	1.49	4.76	21.00	11	3.10	52.38	1.91
2	1.64	9.52	10.50	12	3.15	57.14	1.75
3	1.80	14.29	7.00	13	3.31	61.90	1.62
4	1.90	19.05	5.25	14	3.66	66.67	1.50
5	2.20	23.81	4.20	15	4.01	71.43	1.40
6	2.30	28.57	3.50	16	4.11	76.19	1.31
7	2.50	33.33	3.00	17	4.23	80.95	1.24
8	2.70	38.10	2.63	18	4.50	85.71	1.17
9	2.75	42.86	2.33	19	5.10	90.48	1.11
10	2.80	47.62	2.10	20	5.38	95.24	1.05

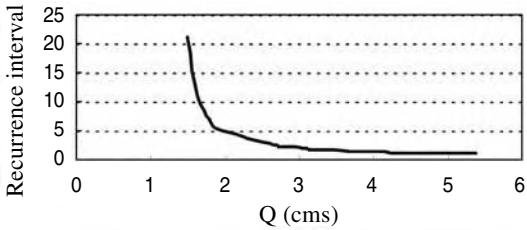


FIGURE 9.3 Frequency curve corresponding to the minimum 7-day flow.

9.2.3.1 QUAL2E Model

The enhanced streamwater quality model (QUAL2E) is a comprehensive and versatile model that can simulate several water quality factors in any combination desired by the user. QUAL2E is a one-dimensional model and uses a finite-difference solution of the mass transport and reaction equation. It allows multiple waste discharges, withdrawals, tributaries, and incremental inflow and outflow. QUAL2E can operate either as a steady-state or as a quasi-dynamic model. Dynamic operation of the model allows study of the effects of diurnal variation of meteorological data on water quality and the effects of diurnal variation of respiration and algal growth on dissolved oxygen. The QUAL2E model considers the major interactions and effects of nutrient cycles, benthic and carbonaceous oxygen demand, atmospheric aeration, and algal production on the dissolved oxygen concentration in a stream. It also can model coliforms and nonconservative as well as conservative constituents that do not decay or interact with other constituents.

QUAL2E-UNCAS is an enhancement to the QUAL2E model that allows users to perform uncertainty analyses on steady-state streamwater quality simulations. Three uncertainty analysis techniques can be employed in this model:

- Sensitivity analysis
- First-order error analysis
- Monte Carlo simulation

For detailed information, the reader is referred to Brown and Barnwell (1987).

9.2.3.2 WASP5

The water quality analysis simulation program (WASP5) is a dynamic compartment modeling framework that can be used to simulate a variety of water quality contaminants in a diverse set of water bodies. The WASP5 has been developed by the EPA's Center for Exposure Assessment Modeling (EPA CEAM). It can be applied to one-, two-, and three-dimensional simulation of transportation and transformation of conventional and toxic pollutants in ponds, streams, lakes, reservoirs, rivers, estuaries, and coastal waters. TOXI5 and EUTRO5 are two major subcomponent models that can be linked to WASP5. TOXI5 is a dynamic model of the transport and fate of organic chemicals and metals in all types of aquatic systems that can simulate

the transport and transformation of up to three chemicals and up to three types of particulate materials. TOXI5 uses a simple sediment balance and organic chemical process kinetics to predict dissolved and sorbed chemical concentrations. EUTRO5 simulates the transport and transformation reactions of a wide range of water quality variables, which can be considered as four interactive systems: phytoplankton kinetics, the phosphorus cycle, the nitrogen cycle, and the dissolved oxygen balance. The transport process can be simulated by coupling this model with external hydrodynamic models (see Ambrose, 1987, for more details).

9.2.4 Simple Streamwater Quality Models

9.2.4.1 Oxygen Sag Model

An oxygen-consuming pollutant, either organic or inorganic, discharged into a stream causes depletion of the dissolved oxygen. In this case, DO is initially consumed faster near the discharge point, which poses a danger for aquatic life if the concentration of oxygen falls below a critical level. Determining the amount of discharged waste and how much oxygen will be required to degrade the waste is necessary for predicting the extent of oxygen depletion.

Oxygen is continuously being replenished from the atmosphere and consumed by organisms, so the oxygen concentration in a river is determined using the relative rates of these competing processes. Downstream of the discharge point, the river recovers and the DO concentration goes back up again. This profile of DO concentration along the river is referred to as *DO sag*. Figure 9.4 shows a plot of DO sag.

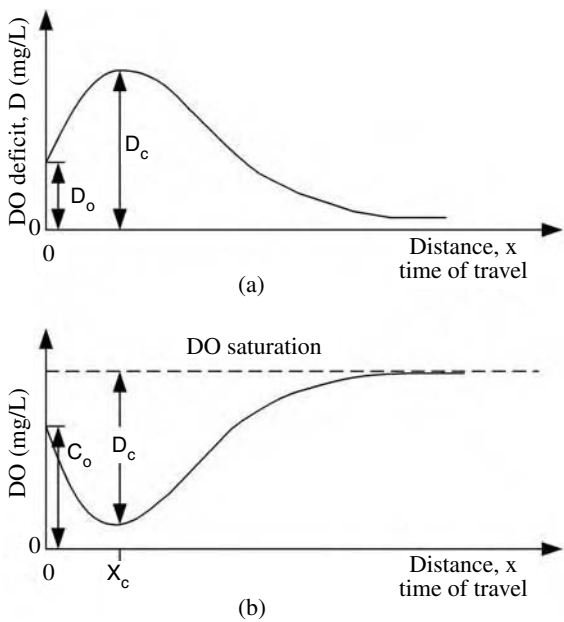


FIGURE 9.4 Dissolved oxygen sag.

The one-dimensional mass balance equation for DO can be derived using Eq. 9.7:

$$\frac{Dc}{Dt} = \frac{\partial}{\partial x}\left(E_{xx}.\frac{\partial c}{\partial x}\right) + k_2\left(O_s - c\right) + P - R - k_c.L_c - k_c.L_n + G_o \tag{9.30}$$

where:

G_o is the distributed dissolved oxygen source.
c is the dissolved oxygen concentration.
k_2 is the reaeration coefficient.
O_s is the dissolved oxygen saturation.
P is the rate of photosynthesis.
R is the rate of respiration.
k_c is the carbonaceous biochemical oxygen demand (CBOD) deoxygenation coefficient.
L_c is the concentration of CBOD.
k_n is the nitrogenous deoxygenation coefficient.
L_n is the nitrogenous biochemical oxygen demand (NBOD) concentration.
E_{xx} is dispersion coefficient in the x direction.

As originally derived by Streeter and Phelps in 1925, if the effects of dispersion, P, R, NBOD, and G_o are neglected, then Eq. (9.30) could be simplified as:

$$\frac{Dc}{Dt} = k_2(O_s - c) - k_c.L_c \tag{9.31}$$

$$L_c = L_{c0}.e^{-k_c^{'}.t} \tag{9.32}$$

where $k_c^{'} = k_c + k_s$, k_s is the rate of settling of CBOD, and L_{c0} is L_c at $t = 0$. Using Eqs. (9.31) and (9.32) and defining the DO deficit as $D = O_s - c$, $D = D_0$ at $t = 0$, $D = 0$ at $t = \infty$, and $k_c^{'} = k_c$, the Streeter–Phelps equation is derived as:

$$D = \frac{k_c.L_{c0}}{k_2 - k_c}\left(e^{-k_c.t} - e^{-k_2.t}\right) + D_0.e^{-k_2.t} \tag{9.33}$$

If the deficit calculated from the above equation is greater than the saturation DO, all the dissolved oxygen has been depleted and the DO is zero. The lowest point on the DO sag curve indicates the worst water quality condition and is of major interest. The time to this minimum DO point, the critical time (t_c), can be found by differentiating Eq. (9.33), setting it equal to 0, and solving for t:

$$t_c = \frac{1}{k_2 - k_c}\ln\left[\frac{k_2}{k_c}\left(1 - D_0 \times \frac{k_2 - k_c}{k_c - L_{c0}}\right)\right] \tag{9.34}$$

TABLE 9.8
Typical Values for the BOD Rate Constant

Sample	k (20°C) (day^{-1})
Raw sewage	0.15–0.30
Well-treated sewage	0.05–0.10
Polluted river water	0.05–0.10

The critical deficit (D_c) is then found by using t_c in Eq. (9.33). The bio-oxidation rate (k_c) is dependent on the waste characteristics, the ability of organisms to oxidize the waste, and the water temperature. Temperature increases the speed of most of the biological processes. Laboratory testing is usually done at a standard temperature of 20°C, and the BOD rate constant at any temperature T (°C) is:

$$k_{c,T} = k_{c,20}.(\theta)^{T-20} \tag{9.35}$$

where $k_{c,T}$ is the BOD rate constant (expressed in 1/day) at temperature T (°C), $k_{c,20}$ is the BOD rate constant determined under laboratory conditions at 20°C (1/day), and θ is the temperature coefficient. The value of θ is considered to be 1.135 for $4 < T < 20$°C and 1.056 for $20 < T < 30$°C. Typical values for the BOD rate constant under laboratory conditions are presented in Table 9.8.

The deoxygenation rate constant (k_c) in a river differs from the BOD rate constant (k) due to physical and biological differences between a bottle and a river. In river systems, the BOD is exerted more rapidly because of turbulent mixing and BOD removal by suspended and sediment organisms. k_c can be estimated from the BOD rate constant (k) by the following equation:

$$k_c = k + \frac{v}{h}\eta \tag{9.36}$$

where:

k_c is the deoxygenation rate constant at 20°C (1/day).
v is the average velocity of streamflow (m/sec).
k is the BOD rate constant determined in laboratory conditions at 20°C (1/day).
η is the bed activity coefficient (0.1 for deep water; 0.6 or greater for a rapidly flowing stream).
h is the average depth (m).

When the stream temperature is not 20°C, the calculated k_c from the above equation should be adjusted using Eq. (9.35). The value of the reaeration rate constant (k_2) depends on the degree of turbulent mixing, which is related to river flow velocity

TABLE 9.9
Range of Reaeration Constants for Various Water Bodies

Water Body	Ranges of k_2 at 20°C (1/day)
Small ponds	0.05–0.1
Small streams of low velocity	0.10–0.15
Large lakes	0.10–0.15
Large streams	0.15–0.30
Swift streams	>0.30

Note: For other temperatures, use $k_{2,T} = k_{2,20}1.024^{T-20}$.

and the surface and depth of the water body (surface area exposed to the atmosphere per unit volume of water). The reaeration rate constant is also related to temperature and should be adjusted to river temperature using Eq. (9.35) but with a temperature coefficient θ, which is equal to 1.024.

Some researchers have developed formulas for calculating k_2 directly from stream physical parameters. One of these formulas was developed by O'Connor and Dobbins (1958) as follows:

$$k_2 = \frac{3.9 \times v^{1/2}}{h^{3/2}} \tag{9.37}$$

The Langbein and Duram (1967) formula can be used to calculate k_2 as follows:

$$k_2 = \frac{2.208 \times v}{h^{4/3}} \tag{9.38}$$

where:

k_2 is the reaeration rate constant at 20°C (1/day).
v is the average stream velocity (m/sec).
h is the average depth (m).

Table 9.9 shows the typical reaeration constants for various surface water bodies.

Example 9.7

A treatment plant discharges 0.15 m^3/sec of partially treated sewage into a free-flowing stream. The stream characteristics upstream of the point of discharge at summer low flow conditions are as follows:

Flow discharge = 0.5 m^3/sec.
Temperature = 22°C.
BOD_5 = 3 mg/l.
DO = 8 mg/l.
Velocity = 0.1 m/sec.
Average depth = 2.2 m.
Bed activity coefficient = 0.2.

The effluent characteristics have been measured as follows:

Temperature = 25°C.
BOD_5 = 40 mg/l.
DO = 2 mg/l
k_2 at 20°C = 0.4 1/day.
k (laboratory test) = 0.22 1/day.

Calculate the DO sag curve for a section located 120 km downstream of the discharge point.

Solution: The characteristics of mixed flow can be computed as:

$$Q_{mix} = 0.15 + 0.5 = 0.65 \frac{\text{m}^3}{\text{s}}$$

$$BOD_{mix} = \frac{(0.5 \times 3) + (0.15 \times 40)}{0.65} = 11.54 \quad \text{mg / L}$$

$$DO_{mix} = \frac{(0.5 \times 8.0) + (0.15 \times 2.0)}{0.65} = 6.62 \quad \text{mg / L}$$

$$T_{mix} = \frac{(0.5 \times 22) + (0.15 \times 25)}{0.65} = 22.7 \quad \text{mg / L}$$

The deoxygenation and reaeration coefficients are calculated using Eqs. (9.36) and (9.37), respectively, as follows:

$$k_{c,20} = 0.22 + \frac{0.1 \times 0.2}{2.2} = 0.23 \frac{1}{\text{day}} \text{ at } 20^{\circ}\text{C}$$

$$k_{2,20} = \frac{3.9 \times (0.1)^{0.5}}{(2.2)^{\frac{3}{2}}} = 0.38 \frac{1}{\text{day}} \text{ at } 20^{\circ}\text{C}$$

k_c and k_2 should be adjusted to 22.7°C as follows:

$$k_{c,22.7} = k_{c,20}(1.056^{22.7-20}) = 0.23 \times 1.158 = 0.266 \frac{1}{\text{day}}$$

$$k_{2,22.7} = k_{2,20}(1.024^{22.7-20}) = 0.38 \times 1.066 = 0.41 \frac{1}{\text{day}}$$

The ultimate BOD (L_{c0}) can be calculated using the following equation:

$$\begin{aligned} BOD_5 &= L_{c0} - L_{c0}.e^{-k_{c,20} \times 5} \\ L_{c0} &= \frac{BOD_5}{1 - e^{-k_{c,20} \times 5}} = \frac{11.54}{1 - e^{-0.23 \times 5}} = 16.88 \text{ mg/L} \end{aligned} \tag{9.39}$$

As the DO saturation (O_s) at 22.7°C is 8.7 mg/L, the initial DO deficit (D_o) can be computed as follows:

$$D_o = 8.7 - 6.62 = 2.08 \text{ mg/L}$$

Now we have all the information we need to calculate the time to the critical point and maximum DO deficit. Using Eqs. (9.33) and (9.34), it can be written that:

$$t_c = \frac{1}{0.41 - 0.26} \ln \left[\frac{0.41}{0.26} \left(1 - 2.08 \times \frac{0.41 - 0.26}{0.26 \times 16.88} \right) \right]$$

$$t_c = 2.54 \quad \text{day}$$

$$D_c = \frac{0.266 \times 16.88}{0.41 - 0.266} \times \left(e^{-0.266 \times 2.54} - e^{-0.41 \times 2.54} \right) + 2.08 \times e^{-0.41 \times 2.54} = 3.56 \text{ mg/L}$$

This critical condition occurs at the following distance downstream of the discharge point:

$$x_c = 0.1 \times 2.54 \times 86,400 = 21,946 \text{ m} = 21.9 \text{ km}$$

By the same method, oxygen deficits at 20, 50, 75, 100, and 120 km downstream of the point of discharge can be calculated as follows:

$$t \text{ (day)} = \frac{x \text{ (m)}}{86,400 \times v \left(\frac{\text{m}}{\text{s}} \right)}$$

$$t_{20} = 2.315 \text{ day} \qquad t_{50} = 5.787 \text{ day}$$

$$t_{75} = 8.680 \text{ day} \qquad t_{100} = 11.57 \text{ day}$$

$$t_{120} = 13.88 \text{ day}$$

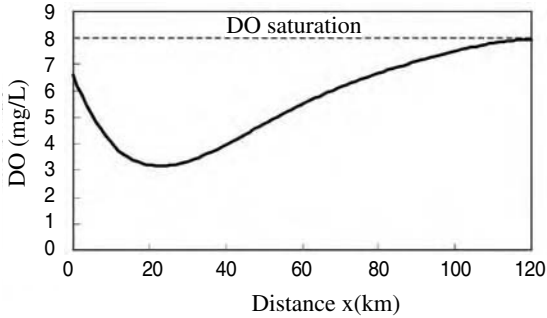


FIGURE 9.5 Dissolved oxygen sag for Example 9.7.

The DO deficit at these times are calculated using Eq. (9.33) as follows:

$$D_{20} = \frac{0.26 \times 16.88}{0.41 - 0.26}\left(e^{-0.26 \times 2.315} - e^{-0.41 \times 2.315}\right) + 2.08e^{-0.41 \times 2.315}$$

$$D_{20} = 5.51 \text{ mg / L}$$

$$D_{50} = 3.96 \text{ mg / L} \qquad D_{75} = 2.29 \text{ mg / L}$$

$$D_{100} = 1.21 \text{ mg / L} \qquad D_{120} = 0.7 \text{ mg / L}$$

The dissolved oxygen concentrations at these points are also computed as:

$$c_{20} = 3.19 \text{ mg / L} \qquad c_{21.6} = 3.17 \text{ mg / L}$$

$$c_{50} = 4.74 \text{ mg / L} \qquad c_{75} = 6.41 \text{ mg / L}$$

$$c_{100} = 7.49 \text{ mg / L} \qquad c_{120} = 8 \text{ mg / L}$$

Figure 9.5 shows the DO sag curve based on these computations.

9.2.4.2 Simple Water Quality Model for Multiple-Discharge Streams

The Streeter–Phelps equation can be used for water quality modeling of streams having several reaches with different flows, reaeration and bio-oxidation parameters, and pollutants discharged from point sources. A typical section of river is presented in Figure 9.6. For an important inflow and withdrawal or where sections of river have different velocities and reaeration coefficients, a separate reach is considered. Each reach is indexed by i and each discharge by j. Each discharge can have an associated withdrawal that changes the stream flow. Each reach has three important points. The first point ($t_{i,0}$), located downstream of discharge point j, is the first boundary condition of reach i. The next important point in each reach is the critical point ($t_{i,c}$) having the greatest oxygen deficit or minimum oxygen concentration. The third important point is located just before the next discharge point. The pollutant

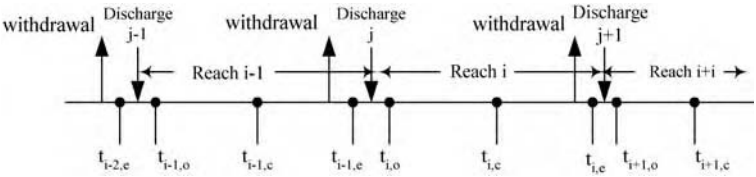


FIGURE 9.6 Schematic diagram of multireach/multidischarge stream model. (From ReVelle, C. and McGarity, A. E., *Design and Operation of Civil and Environmental Engineering Systems*, John Wiley & Sons, New York, 1997. With permission.)

concentration is calculated at points $t_{i,0}$ and $t_{i,e}$ to evaluate dilution effects at the discharge points. The pollutant concentration at critical point $t_{i,c}$ is compared with the water quality standard to determine whether any water quality standards are being violated.

Assume that Q_i is the flow rate in reach i (m^3/sec), $q_{j\text{-}out}$ is the withdrawal flow rate at point j (m^3/sec), and $q_{j\text{-}in}$ is the discharged flow rate of discharge point j. It should be noted that in this section only the CBOD is considered; therefore, notation L is used instead of L_c. For BOD pollutants and considering complete mixing at discharge points, we have.

$$L_{i,0} = \frac{Q_{i-1} - q_{j,out}}{Q_i}.L_{i-1,e} + \frac{q_{j,in}}{Q_i} l_j \tag{9.40}$$

$$D_{i,0} = \frac{Q_{i-1} - q_{j,out}}{Q_i}.D_{i-1,e} + \frac{q_{j,in}}{Q_i} d_j \tag{9.41}$$

where:

$L_{i,s}$ and $D_{i,s}$ are the BOD concentration and DO deficit, respectively, at point s in reach i.
l_j is the BOD concentration of discharge j (mg/l).
d_j is the DO deficit of discharge j (mg/l).

The BOD concentration at points $s = c$ and e on reach i can be calculated using the BOD decay equation:

$$L_{i,s} = L_{i,0} \times e^{-k_{ci} \cdot t_{i,s}} \tag{9.42}$$

And, from the Streeter–Phelps equation, the DO deficit at points $s = c$ and e in reach i are:

$$D_{i,s} = \frac{k_{ci}.L_{i,0}}{k_{2i} - k_{ci}}.\left(e^{-k_{ci} \cdot t_{i,s}} - e^{-k_{2i} t_{i,s}}\right) + D_{i,0}.e^{-k_{2i} \cdot t_{i,s}} \tag{9.43}$$

The critical DO deficit in reach i can be derived from Eqs. (9.31) and (9.32) as follows:

$$D_{i,c} = \frac{k_{ci}}{k_{2i}} \cdot L_{i,0} \cdot e^{-k_{ci} \cdot t_{i,c}} \tag{9.44}$$

The time t at each point can be calculated based on the distance the pollutant has traveled divided by the velocity ($t = x/v$), where x is the distance traveled and v is the average streamflow velocity. Therefore, when considering the average velocity in each reach, Eqs. (9.42) and (9.43) are simplified to linear equations and can easily provide dissolved oxygen and BOD concentrations at each point of the stream.

9.2.5 Water Quality Management Models for Rivers and Streams

In this section, we describe deterministic and stochastic models that have been developed for river and streamwater quality management and waste load allocation. Linear, nonlinear, and dynamic programming techniques have been used to develop optimal management policies and to identify optimal solutions for waste load allocation problems. The study of waste load allocation as an important element of river water quality management can be necessary in several situations. The most common problem arises when the amount of pollutant removal should be determined at a number of discharge points along a water body in order to achieve or maintain an acceptable level of water quality. Another situation could arise from the capacity expansion problem, where in one or more point sources influent loads are increased. In these cases, the appropriate increase in the size and capacity of treatment facilities must be determined. Another situation may occur when an additional discharger wishes to locate on a water body that could necessitate reallocation of the assimilative capacity of the water body among the existing discharges (Burn, 1989). Minimization of treatment costs has been considered by various investigators as the objective function of water quality management models developed for streams. An ideal strategy can be obtained by the combined use of optimization and simulation models, but the optimization process can be more difficult.

9.2.5.1 Deterministic Approach

9.2.5.1.1 Linear Programming Models for Dissolved Oxygen Management

In this section, a management model is developed based on the multireach/multidischarge system presented in the previous section. The objective function of this model is minimization of treatment costs. The general form of the wastewater treatment cost function (see Figure 9.7) is a nonlinear function of the treatment efficiency. The cost curve shows three classes of treatment: primary, secondary, and tertiary. According to Figure 9.7, the treatment costs can be represented by the cost function $c(x)$, where x is the treatment plant efficiency. The cost function has both

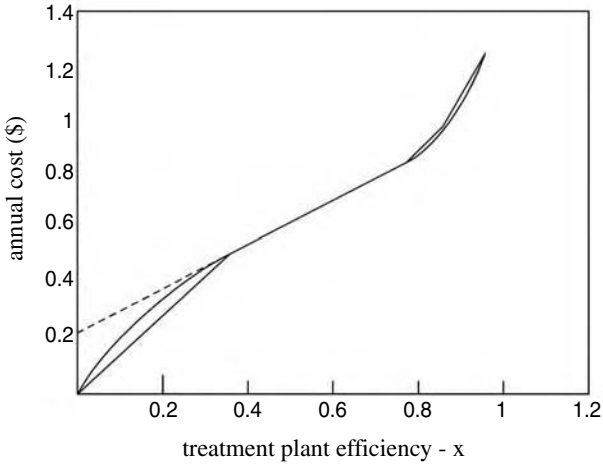


FIGURE 9.7 Annual costs of treatment vs. treatment efficiency. (From ReVelle, C. and McGarity, A. E., *Design and Operation of Civil and Environmental Engineering Systems*, John Wiley & Sons, New York, 1997. With permission.)

convex and concave sections. By assuming that the treatment efficiency is greater than the primary treatment efficiency, the cost function can be changed to convex form by linearly extending the secondary treatment section to the origin (dashed line in Figure 9.7). The linear programming technique can also be used; for this purpose, piecewise linear approximation of the cost function should be used. The total cost of wastewater treatment can be written as follows (ReVelle and McGarity, 1997):

$$\text{Total Cost} = \sum_{j=1}^{m} (c_{j0} + c_{j1}.x_{j1} + c_{j2}.x_{j2} + c_{j3}.x_{j3}) \tag{9.45}$$

where m is the number of treatment plants; c_{j0} is the intercept of the convex approximation with the cost axis ($/year); c_{j1}, c_{j2}, and c_{j3} are the slopes of the linear approximation for the secondary, first tertiary, and second tertiary treatment segments, respectively ($/year/efficiency fraction); x_{j1}, x_{j2}, and x_{j3} are the efficiency segments in the secondary, first tertiary, and second tertiary treatment segments, respectively; and $x_j = x_{j1} + x_{j2} + x_{j3}$.

In general, Eq. (9.45) can be written as:

$$Z = \sum_{j=1}^{m} (c_{j0} + \sum_{k=1}^{P_j} c_{jk}.x_{jk}) \tag{9.46}$$

where P_j is the number of linear segments that are required to present the treatment cost function. The relationship between l_j^u and l_j, the BOD concentration in untreated and treated wastewater discharge j, respectively, can by presented as:

$$x_j = \frac{l_j^u - l_j}{l_j^u} \qquad j = 1,...,m \tag{9.47}$$

where l_j^u and l_j are expressed in mg/L, and x_j is the efficiency of BOD removal. As described before, the BOD concentration and oxygen deficit at the end of each reach can be expressed as:

$$L_{i,e} = \rho_{i,e}.L_{i,0} \qquad i = 1,2,...,m \tag{9.48}$$

$$D_{i,e} = \alpha_{i,e}.L_{i,0} + \beta_{i,e}.D_{i,0} \qquad i = 1,2,...,m \tag{9.49}$$

Using the average velocity of stream flow, the $t_{i,e}$ can be calculated; therefore, $\rho^i_{i,e}$, $\alpha^i_{i,e}$, and $\beta^i_{i,e}$ are constants for discharger $j = i$ into reach i. The DO deficit at point s in each reach i can be estimated as:

$$D_{i,s} = \alpha_{i,s}L_{i,o} + \beta_{i,s}.D_{i,o} \qquad (i = l,\ldots,m) \tag{9.50}$$

In order to estimate where the critical points are, each reach i is divided into n_i segments; the DO deficits are calculated at the end of each segment using the flow travel times $t_{i,s}$ (i = 1, …, m; s = 1, …, n_i) and the parameters $\alpha_{i,s}$ and $\beta_{i,s}$. The DO deficit is limited to maximum allowable deficit at each reach i ($D_{i,max}$). The number of segments at each reach (n_i) can be calculated based on the reach length and characteristics of the oxygen sag curve. Following is the complete linear formulation for DO and BOD concentration control in streams using treatment cost minimization as the objective function (ReVelle and McGarity, 1997):

Minimize $$z = \sum_{j=1}^{n}\sum_{k=1}^{P_j} c_{jk}x_{jk} \tag{9.51}$$

Subject to:

$$\sum_{k=1}^{P_j} x_{jk} + \frac{1}{l_j^u} l_j = 1 \qquad j = 1,2,...,m \tag{9.52}$$

$$\eta_{j,\min} \le x_{j1} \le \eta_{j1} \qquad j = 1,2,...,m \tag{9.53}$$

$$0 \le x_{jk} \le (\eta_{jk} - \eta_{j,k-1}) \qquad \begin{matrix} k = 2,3,...,P_j \\ j = 1,2,...,m \end{matrix} \tag{9.54}$$

$$(Q_{i-1} - q_{j,out})L_{i-1,e} + q_{j,in}l_j - Q_i L_{i,0} = 0 \qquad \begin{matrix} i = 1,2,...,m \\ j = i \end{matrix} \tag{9.55}$$

$$(Q_{i-1} - q_{j,out})D_{i-1,e} + q_{j,in}d_j - Q_i D_{i,0} = 0 \qquad \begin{matrix} i = 1,2,...,m \\ j = i \end{matrix} \tag{9.56}$$

$$L_{i,e} - \rho_{i,e}L_{i,0} = 0 \qquad i = 1,2,...,m \tag{9.57}$$

$$D_{i,e} - \alpha_{i,e}L_{i,0} - \beta_{i,e}D_{i,0} = 0 \qquad i = 1,2,...,m \tag{9.58}$$

$$\alpha_{i,s}L_{i,0} - \beta_{i,s}D_{i,0} \le D_{i,\max} \qquad \begin{matrix} i = 1,2,...,m \\ s = 1,2,...,n_i \end{matrix} \tag{9.59}$$

$$D_{i,0} \le D_{i,\max} \qquad i = 1,2,...,m \tag{9.60}$$

$$l_j \ge 0 \qquad j = 1,2,...,m \tag{9.61}$$

$$L_{i,e} \ge 0 \qquad i = 1,2,...,m \tag{9.62}$$

where:

$\eta_{j,min}$ is the lowest treatment efficiency allowed at wastewater discharge j.
P_j is the number of linear segments necessary to represent the treatment cost curve for wastewater discharge j.
$\eta_{j,k-1}$ and $\eta_{j,k}$ are the upper and lower limits, respectively, of the kth segment.
$q_{j,in}$ and $q_{j,out}$ are the volumetric flow rates of discharge and withdrawal, respectively, at point j.
l_j and d_j are the BOD and DO concentrations, respectively, of wastewater discharge j.

9.2.5.1.2 General Cost-Minimization Model for Water Quality Management

In the previous section, a cost-minimization model was presented for dissolved oxygen management in rivers and streams. This model can be extended for an arbitrary water quality parameter as follows (for a detailed description, see ReVelle and McGarity, 1997):

$$\text{Minimize} \quad Z = \sum_{j=1}^{m} C_j(x_j) \tag{9.63}$$

$$\text{Subject to:} \quad \sum_{j=1}^{m} a_{i,j} x_j \geq b_i \qquad i = 1,2,...,n \tag{9.64}$$

$$l_j < x_j < u_j \qquad j = 1,2,...,\text{m} \tag{9.65}$$

where:

x_j is the treatment efficiency at discharge j.
$C_j(x_j)$ is the treatment cost function at discharge j.
m is the number of dischargers on the stream.
$a_{i,j}$ is the transfer coefficient between the pollutant source or discharger j and checkpoint i indicating water quality improvement at point i as a result of pollution removal at discharge j.
n is the number of checkpoints in the stream.
b_i is the required water quality improvement at checkpoint i to satisfy water quality standards.
l_j and u_j are the lower and upper limits, respectively, of the treatment efficiency of discharge j.

Example 9.8

Many small communities and some small industries are located on a river. In a stretch of the river two major polluters are located: Industry A at river mile 310 and City B at river mile 302 with some industries. For this example, it can be assumed that the water quality of the river is important from river mile 310 to river mile 275 and is particularly important at river mile 294, where a lot of swimming and boating take place. The requirement at river mile 294 is that the dissolved oxygen concentration should be equal to or exceed 5 mg/L. At other points between river miles 310 and 275, the dissolved oxygen concentration must be not less than 4 mg/L. Both Industry A and City B already have some treatment facilities, so their treatment efficiencies already exceed 30% (i.e., x_1, $x_3 \geq .30$). In addition, due to the nature of the wastes and current treatment practices, it would be impossible to achieve greater than 95% efficiencies (i.e., x_1, $x_3 \leq .95$). River C is a small tributary and has no treatment plant associated with it ($x_2 = 0$).

The data provided in Table 9.10 are from laboratory analyses of water just below river miles 310, 307, and 302. The water samples were enclosed in airtight bottles, and the BOD_5 concentrations of the water samples were measured every 3 hours. The water was collected on a summer day when the streamflows were equal to the 7-day, 10-year low flow; the water temperature was kept at 20°C during the analysis. The flow rate at river mile 310.1 (upstream of Industry A discharger) was measured as 3 (m^3/sec). The results are presented in Table 9.10.

TABLE 9.10
Results of Measurement of BOD_5 Concentrations (Example 9.8)

Time (hr)	At Mile 309.9 (mg/l)	At Mile 306.9 (mg/l)	At Mile 301.9 (mg/l)
0	13.4	11.4	16.7
3	13.1	11.1	16.3
6	12.7	10.8	15.9
9	12.4	10.6	15.5
12	12.1	10.3	15.1
15	11.8	10.1	14.7
18	11.5	9.8	14.4
21	11.2	9.6	14.0
24	11.0	9.3	13.7

TABLE 9.11
Qualitative and Quantitative Characteristics of River and Wastewater Discharge

Reach or Discharge Point	Streamflow Velocity (m/day)	Discharge Rate (m^3/sec)	Influent BOD_5 (mg/L)	Influent DO (mg/L)	Average Stream Depth (m)
1	16062.96	0.0566	1020	8	0.732
2	10789.92	0.339	3	7	0.543
3	12923.52	0.283	210	7	0.564

1. For each reach of the river, estimate the value of the bio-oxidation rate constant (k_c) from the data provided. Some field measurements were made last summer during what was considered to be the 7-day, 10-year low flow for this river. The collected data are presented in Table 9.11.
2. For each reach of the river, estimate the value of the reaeration rate constant (k_2) using the Langbein–Duram formula.

   The quantitative and qualitative characteristics of wastewater discharge points are given in Table 9.11 and will be assumed to be constant in the remainder of this example. Last summer, the data in Table 9.12 were collected at the same time as the data of Table 9.11. The treatment efficiencies of both Industry A and City B were 50%. The actual BOD_5 concentration in the treated effluents of Industry A and City B were 510 mg/l and 105 mg/l, respectively. The water temperature was 27°C throughout the river, and the DO saturation concentration was 8.0 mg/l.
3. Use the values of k_c and k_2 determined in parts 1 and 2 and the Streeter–Phelps equation to simulate the BOD and DO for river miles 301 to 280. Be sure to adjust the values of k_c and k_2 for changes in temperature (the temperature coefficient, θ, is considered to be 1.047 and 1.016 for k_c and k_2, respectively). You should find that, between river miles 310 and 302, the estimated BOD and DO values closely match the data of

TABLE 9.12
BOD and DO Concentration (mg/l) along the River (Example 9.8)

River Mile (1)	BOD_5 (Measured) (2)	BOD_5 (Computed) (3)	BOD_5 (Computed with Correction) (4)	DO (Measured) (5)	DO (Computed) (6)	DO (Computed with Correction) (7)
310.2	4.0	—	—	7.0	—	—
309.9	13.4	13.33	—	7.0	6.99	—
309.0	13.0	13.0	—	6.7	6.73	—
308.0	12.7	12.46	—	6.5	6.47	—
307.2	12.3	12.32	—	6.3	6.26	—
306.9	11.4	11.32	—	6.3	6.28	—
306.0	11.0	10.90	—	6.1	6.05	—
305.0	10.5	10.46	—	5.9	5.82	—
304.0	10.1	10.04	—	5.7	5.64	—
303.0	9.7	9.64	—	5.5	5.49	—
302.2	9.3	9.28	—	5.4	5.38	—
301.9	16.7	16.56	16.45	5.5	5.46	5.39
301.0	15.5	16.05	15.42	4.6	5.19	4.59
300.0	14.4	15.51	14.31	3.9	4.96	3.86
296.0	10.7	13.52	10.61	2.2	4.31	2.19
292.0	7.9	11.79	7.87	1.9	4.10	1.86
288.0	5.9	10.28	5.84	2.2	4.13	2.21
284.0	4.4	8.96	4.33	2.9	4.32	2.87
280.0	3.2	7.81	3.21	3.7	4.58	3.64

Table 9.12. Between river miles 301 and 280, the simulated BOD and DO should vary significantly from the data presented in Table 9.12.

4. Assume that the value of k_c for river miles 301 to 275 equals the value found in part 2. Estimate a new value of k_2 for river miles 301 to 275 that reproduces the BOD and DO data of Table 9.12.
5. By solution of a linear program, find the minimum percentage removal for Industry A and City B (i.e., the treatment efficiencies of both polluters are equal) that ensures a DO concentration of at least 4 mg/L along the river.

Solution:

1. The bio-oxidation rate constant for each station can be calculated using Eq. (9.32) as follows:

$$L_c = L_{c0} \times e^{-k_c \cdot t}, \frac{L_c}{L_{co}} = e^{-k_c \cdot t}, \text{ and } k_c = \frac{-\ln\left(\frac{L_c}{L_{c0}}\right)}{t}$$

TABLE 9.13
Values of $-\ln(L_c/L_{c0})$ for Stations

Time (hr)	Mile 309.9 (mg/L)	Mile 306.9 (mg/L)	Mile 301.9 (mg/L)
0	—	—	—
3	0.0226	0.0267	0.0242
6	0.0537	0.0541	0.0491
9	0.0776	0.0728	0.0746
12	0.1020	0.1015	0.1007
15	0.1272	0.1211	0.1276
18	0.1529	0.1512	0.1482
21	0.1793	0.1719	0.1764
24	0.1974	0.2036	0.1980

The values of $-\ln(L_c/L_{c0})$ are presented in Table 9.13. If we draw the value of $-\ln(L_c/L_{c0})$ vs. t, the slope of this line will give the value of k_c. Using this method, the value of k_c and the regression coefficient are calculated for each station as follows:

$$x = 309.9 \text{ mile:} \quad k_{c1} = 0.2023 \quad R = 0.99$$
$$x = 306.9 \text{ mile:} \quad k_{c2} = 0.2002 \quad R = 0.99$$
$$x = 301.9 \text{ mile:} \quad k_{c3} = 0.1998 \quad R = 0.99$$

The values of the regression coefficients (near 1) show that the values of k_c are acceptable.

2. The values of k_2, in different reaches using Eq. (9.38), can be calculated as:

$$\text{Reach 1:} \quad k_2 = \frac{2.2208 \times 16062.96}{86400(0.732)^{\frac{4}{3}}} = 0.626$$

$$\text{Reach 2:} \quad k_2 = 0.626$$

$$\text{Reach 3:} \quad k_2 = 0.713$$

3. The coefficients of k_c and k_2 should be adjusted using Eq. (9.35) as follows:

Reach 1:

$$k_{c,27} = 0.2023(1.047)^{(27-20)} = 0.2790$$

$$k_{2,27} = 0.626(1.016)^{(27-20)} = 0.7$$

Reach 2:

$$k_{c,27} = 0.2761 \quad \text{and} \quad k_{2,27} = 0.7$$

Reach 3:

$$k_{c,27} = 0.2756 \quad \text{and} \quad k_{2,27} = 0.797$$

It is assumed that correction of k_c based on Eq. (9.36) is negligible. Considering the treatment efficiency of each treatment plant, the BOD concentration of each wastewater discharge point is estimated as follows:

$$j = 1\text{: } x_1 = 50\%, \text{ and } l_1 = 0.5 \times 1020 = 510 \text{ mg / L}$$
$$j = 2\text{: } x_2 = 0\%, \text{ and } l_1 = 3 \text{ mg / L}$$
$$j = 3\text{: } x_1 = 5 = 50\%, \text{ and } l_1 = 0.5 \times 210 = 105 \text{ mg / L}$$

The mixed BOD and DO concentrations just downstream of discharge point 1 are calculated as:

$$DO_{mix} = \frac{(3 \times 7) + (0.0566 \times 8)}{3.056} = 7.019 \quad \text{mg / L}$$
$$BOD_{mix} = \frac{(3 \times 4) + (0.0566 \times 510)}{3.056} = 13.36 \quad \text{mg / L}$$

The DO and BOD concentrations in each discharge point are calculated using Eqs. (9.42) and (9.43). The travel time ($t_{i,s}$) for each distance (s) from initial point of reach i is computed using reach average velocity. For example, in reach 1, Eqs. (9.42) and (9.43) are used as follows:

BOD concentration at point s:

$$L_{c,s} = L_{c,0} e^{-k_c t_s} = 13.36 e^{-0.279 t_s}$$

DO concentration at point s can be calculated considering $C_s = O_s - D_s$ as follows:

$$C_s = O_s\left(1 - e^{-k_2 t_s}\right) + D_o e^{-k_2 t_s} \frac{k_c L_{c0}}{k_2 - k_c}\left(e^{-k_c t_s} - e^{-k_2 t_s}\right)$$
$$C_s = 8\left(1 - e^{-0.7 t_s}\right) + 7.019 e^{-0.7 t_s} - \left(\frac{0.279 \times 13.36}{0.7 - 0.279}\right) \times \left(e^{-0.279 t_s} - e^{-0.7 t_s}\right) \tag{9.66}$$

At river mile 309.9, $t_s = 0.01$ day and the BOD and DO concentrations are:

$$L = 13.36e^{-0.279\times0.01} = 13.3 \text{ mg/L}$$

$$C = 8\left(1 - e^{-0.7\times0.01}\right) + 7.019e^{-0.7\times0.01} - \left(\frac{0.279\times13.36}{0.7-0.279}\right)\times$$

$$\left(e^{-0.279\times0.01} - e^{-0.7\times0.01}\right) = 6.99 \text{ mg/L}$$

The computed BOD and DO concentrations for different stations along the river are presented in columns 3 and 6 of Table 9.12.

4. The new value of k_2 for reach 3 can be calculated based on the data (column 2 of Table 9.12). Using this method (which was presented in the first part of this example), a new value of k_c for reach 3 is computed as 0.6. The new values for BOD and DO concentrations at stations in reach 3 can be found in columns 4 and 7 of Table 9.12.
5. Linear programming models can be used to calculate optimum treatment efficiencies. The objective function of this model is considered as:

$$\text{Minimize } z = x_1$$

or

$$\text{Minimize } z = x_3$$

where x_1 and x_2 are treatment efficiencies for treatment plant 1 and 3 at river miles 310.0 and 302.0. The constraints of this optimization model are as follows:

Constraints related to treatment efficiencies

$x_1 = x_3$, $x_2 = 0$, and $0.3 \le x_1, x_3 \le 0.95$

$l_1 = 1020(1 - x_1)$

$l_2 = 3(1 - x_2)$

$l_3 = 210(1 - x_3)$

Constraints related to BOD and DO concentrations at the initial and endpoints of each reach using Eqs. (9.32), (9.40), (9.41), and (9.66)

$C_{1,0} = 7.019$

$L_{1,0} = 3.926 + 0.0185\ L_1$

$C_{1,e} = 7.2048 - 0.0723\ L_{1,0}$

$L_{1,e} = 0.93\ L_{1,0}$

where $C_{1,0}$ and $C_{1,e}$ are the DO concentrations and $L_{1,0}$ and $L_{1,e}$ are the BOD concentrations at the initial and endpoints of reach 1, respectively.

$C_{2,0} = 0.9\ C_{1,e} + 0.7$

$L_{2,0} = 0.9\ L_{1,e} + 0.3$

TABLE 9.14
BOD and DO Concentrations at River Checkpoints (Example 9.8)

River Mile	Travel Time from River Mile 310.0	DO Concentration (mg/L)	BOD Concentration (mg/L)
309.9	0.010	$DO_1 = 0.0558 + 6.970 - 0.00276\ L_1$	$BOD_1 = 0.997\ L_1$
307.1	0.290	$DO_2 = 1.470 + 5.728 - 0.0702\ L_1$	$BOD_2 = 0.922\ L_1$
306.9	0.015	$DO_3 = 0.0836 + 0.990\ C_2 - 0.0041\ L_2$	$BOD_3 = 0.996\ L_2$
302.1	0.731	$DO_4 = 3.204 + 0.599\ C_2 - 0.1418\ L_2$	$BOD_4 = 0.817\ L_2$
300	0.012	$DO_5 = 0.0761 + 0.990\ C_3 - 0.00714\ L_3$	$BOD_5 = 0.993\ L_3$
301.9	0.249	$DO_6 = 1.440 + 0.820\ C_3 - 0.1256\ L_3$	$BOD_6 = 0.861\ L_3$
298	0.498	$DO_7 = 2.621 + 0.672\ C_3 - 0.2111\ L_3$	$BOD_7 = 0.742\ L_3$
296	0.747	$DO_8 = 3.589 + 0.551\ C_3 - 0.2662\ L_3$	$BOD_8 = 0.639\ L_3$
294	0.996	$DO_9 = 4.383 + 0.452\ C_3 - 0.2985\ L_3$	$BOD_9 = 0.550\ L_3$
290	1.494	$DO_{10} = 5.568 + 0.304\ C_3 - 0.3169\ L_3$	$BOD_{10} = 0.408\ L_3$
286	1.992	$DO_{11} = 6.364 + 0.204\ C_3 - 0.2992\ L_3$	$BOD_{11} = 0.303\ L_3$
282	2.490	$DO_{12} = 6.900 + 0.137\ C_3 - 0.2651\ L_3$	$BOD_{12} = 0.224\ L_3$
278	2.988	$DO_{13} = 7.261 + 0.092\ C_3 - 0.2256\ L_3$	$BOD_{13} = 0.166\ L_3$
274	3.486	$DO_{14} = 7.503 + 0.062\ C_3 - 0.1868\ L_3$	$BOD_{14} = 0.123\ L_3$

$$C_{2,e} = -0.1437\ L_{2,0} + 0.593\ C_{2,0} + 3.254$$

$$L_{2,e} = 0.814\ L_{2,0}$$

where $C_{2,0}$ and $C_{2,e}$ are the DO concentrations and $L_{2,0}$ and $L_{2,e}$ are the BOD concentrations at the initial and endpoints of reach 2, respectively.

$$C_{3,0} = 0.923\ C_{2,e} + 0.539$$

$$L_{3,0} = 0.831\ L_{2,e} + 0.0769 L_3$$

$$C_{3,e} = 0.0686\ C_{3,0} - 0.196\ L_{3,0} + 7.451$$

$$L_{3,e} = 0.4\ L_{3,0}$$

where $C_{3,0}$ and $C_{3,e}$ are the DO concentrations and $L_{3,0}$ and $L_{3,e}$ are the BOD concentrations at the initial and endpoints of reach 3, respectively. Other constraints that give the DO and BOD concentrations at other river miles are presented in columns 3 and 4 of Table 9.14. The proposed linear model was solved using a linear programming package. The optimum treatment efficiency for both treatment plants is 70%. The BOD and DO concentrations along the river based on these treatment efficiencies are shown in Figure 9.8.

9.2.5.1.3 Management Model To Control Water Quality Violation

This model minimizes the worst water quality response within the river or stream system. In this formulation, it is assumed that the treatment cost of dischargers may be expressed as piecewise linear cost functions. The model is of the form (Burn and Lence, 1992):

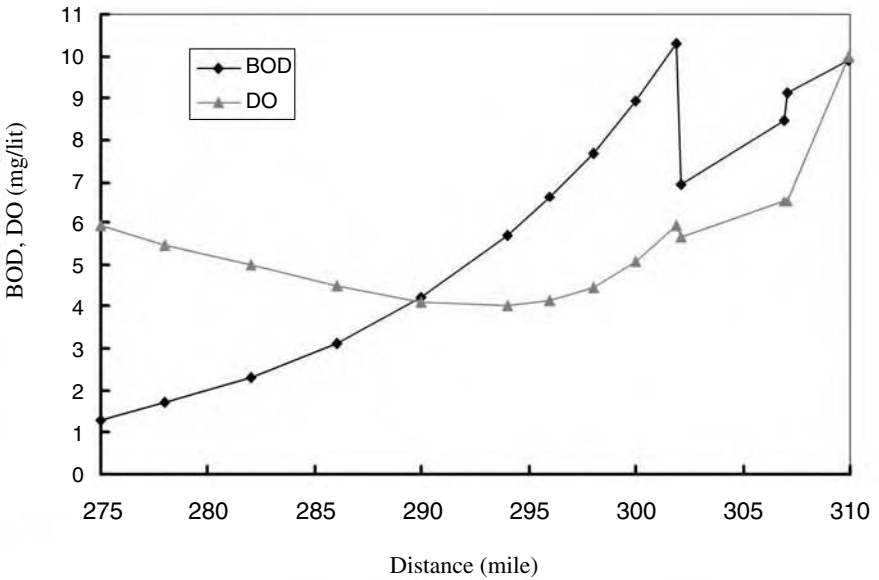


FIGURE 9.8 Dissolved oxygen and BOD concentrations along the river (Example 9.8).

$$\text{Minimize} \quad Z = MV \tag{9.67}$$

$$\text{Subject to:} \quad \sum_j W_j(1-R_j)a_{ij} + BG_i + U_i - V_i = O_{s,i} - DO_i \tag{9.68}$$

$$R_j - \sum_k^{P_j} x_{jk} = R_j^{\min} \qquad j = 1,2,...,m \tag{9.69}$$

$$\sum_{j=1}^{m}\sum_k^{P_j} c_{jk}.x_{jk} \le B \tag{9.70}$$

$$\eta_{j,\min} \le x_{j1} \le \eta_{j,1} \qquad j = 1,2,...,m \tag{9.71}$$

$$0 \le x_{jk} \le \eta_{j,k} - \eta_{j,k-1} \qquad j = 1,2,...,m \tag{9.72}$$

$$MV - V_i \ge 0 \qquad \forall i \tag{9.73}$$

where:

MV is the maximum violation at any checkpoint (mg/L).
W_j is the influent waste load of BOD for discharge point j (kg/day).
R_j is the removal fraction of source j.
a_{ij} is the transfer coefficient between source j and checkpoint i (milligram per liter of DO deficit per kilogram per day of BOD loading).
BG_i is the background DO deficit concentration at checkpoint i.
U_i is the amount by which the DO concentration exceeds the standard level at checkpoint i (mg/L).
V_i is the violation of a water quality standard at checkpoint i (mg/L).
Cs_i is the saturation DO concentration at checkpoint i.
DO_i is the DO standard at checkpoint i.
x_{jk} is the removal fraction at source j, which is associated with treatment on the kth linear segment of piecewise linear cost function.
R_j^{min} is the minimum allowable removal fraction at discharge j.
c_{jk} is the slope of the kth linear segment of the treatment cost curve for discharge j.
B is the total budget for treatment ($/yr).
P_j is the number of linear segments necessary to represent the treatment cost curve for wastewater discharge j.

The violation depends on the treatment efficiency at the sources and the transfer coefficients and background dissolved oxygen deficits. Equation (9.68) defines the violation of the DO standard at location i, Eq. (9.70) constrains the amount of money required for treatment to be less than the budgeted amount, and Eq. (9.73) states that the maximum violation is equal to or greater than each of the violations.

9.2.5.1.4 Deterministic Dynamic Programming Model for Water Quality Management

One of the important advantages of dynamic programming for water quality management is that it permits the use of more realistic water quality simulation models. In contrast to many water quality simulation models, linearity is inherent in the Streeter–Phelps equation. In this section, a deterministic dynamic programming is presented that minimizes the total treatment cost such that water quality standards are met at boundaries. The recursive function of this model is:

$$v(s|n)_{k=k'} = r^{k'} + v(s^*|n-1) \tag{9.74}$$

where v is treatment cost; $r^{k'}$ is the treatment cost of control option k'; and s^* is the resultant water quality state in stage $n-1$ if control option k' is applied in stage n.

Water quality state s at stage n is mapped onto water quality state s^* at stage $n-1$ by control action k through the transition function T as follows:

$$s^* = T(s,k)$$

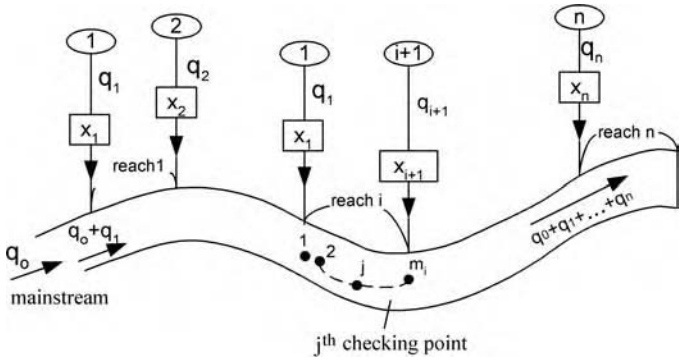


FIGURE 9.9 River reaches and dischargers. (From Fujiwara, O., Puangmaha, W., and Hanaki, K., *ASCE J. Environ. Eng.*, 114(4), 1988. With permission.)

The objective function of the dynamic programming model minimizes the overall cost of all feasible control actions (Cardwell and Ellis, 1993):

$$v(s|n) = \min_{k \in K} (r^k + v(s^*|n-1)) \tag{9.75}$$

The feasible control set K consists of all control options k that give acceptable water quality at the end of the reach:

$$K = [k|s^* \geq S] \tag{9.76}$$

where S is the standard level of water quality variable. One of the problems of deterministic dynamic programming models is that, when a water quality standard must be attained (i.e., treated as an inviolate lower bound), control actions that implicitly or explicitly create transitions to an unacceptable downstream state are declared not feasible (Cardwell and Ellis, 1993).

9.2.5.2 Stochastic Approach

Streamwater quality management is complex because the rivers and streams are polluted from multiple sources, and water quality depends on uncertain streamflow conditions. Therefore, it is necessary to use optimization models that incorporate both uncertainty and complex characteristics of the pollution problem. Stochastic optimization, in which streamflow and/or wastewater flow rate and also various river parameters are assumed to be random variables, have been formulated by many researchers (e.g., Lohani and Thanh, 1978; Cardwell and Ellis, 1993; Takyi and Lence, 1999). In this section, an optimization model with an ability to reflect important uncertainties of streamwater quality problems is demonstrated. This model, which was developed by Fujiwara et al. (1988), optimizes wastewater treatment efficiencies at source points and considers various random parameters of the

river. Quintiles of random variables concerning DO concentration are calculated using a Monte Carlo simulation.

As shown in Figure 9.9, the river has n pollutant sources, and the distance between two sources is considered as a reach. Each pollutant source has a wastewater treatment plant, and m_i water quality control points are considered at each reach i. The objective function of the model is to minimize the total sum of weighted DO deficits at all checkpoints. Based on the Streeter–Phelps model, the DO deficit at the jth checkpoint of reach i can be defined as follows (Loucks et al., 1967; Fujiwara et al., 1988):

$$D_{ij} = \sum_{k=1}^{i} \left[\left(\frac{a_{ij}^{k}}{\sum_{l=0}^{i} q_l} \right) b_k (1 - x_k) \right] + \left(\frac{d_{ij}}{\sum_{l=0}^{i} q_l} \right) \tag{9.77}$$

where:

DD_{ij} is the DO deficit at the jth checkpoint in reach i.
a_{ij}^{k} and d_{ij} are functions of the deoxygenating and reaeration constants, sedimentation coefficients, BOD additions by resuspension of sediments, oxygen reductions by benthal demand, oxygen productions by photosynthesis, and flow travel time.
q_k is the wastewater discharge at the kth pollutant source.
b_k is the BOD loading before entering the kth treatment plant.
x_k is the wastewater treatment efficiency of the kth discharger.

The objective function of the model that minimizes the weighted sum of DO deficits can be written as follows:

$$\text{Minimize} \quad \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} \left\{ \sum_{k=1}^{i} \left[\left(\frac{a_{ij}^{k}}{\sum_{l=0}^{i} q_l} \right) b_k (1 - x_k) \right] + \frac{d_{ij}}{\sum_{l=0}^{i} q_l} \right\} \tag{9.78}$$

where n is total number of pollutant sources and w_{ij} are the relative weights of D_{ij}.

Equation (9.78) can be rewritten by dropping the terms that are independent of the decision variables (x_k):

$$\text{Maximize} \quad \sum_{k=1}^{n} a_k b_k x_k \tag{9.79}$$

where:

$$a_k = \sum_{i=k}^{n} \sum_{j=1}^{m_i} \left(\frac{w_{ij} a_{ij}^k}{q_0 + q_1 + \cdots + q_k} \right) \tag{9.80}$$

For a piecewise linear BOD removal cost function, the optimization model has the following linear form:

$$\text{Maximize} \quad \boldsymbol{g}^T \boldsymbol{x} \tag{9.81}$$

$$\text{Subject to:} \quad \boldsymbol{B}\boldsymbol{x} \le \boldsymbol{h} \tag{9.82}$$

$$\boldsymbol{x} \ge \boldsymbol{0} \tag{9.83}$$

where:

$\boldsymbol{g}$ is a column vector of n random components ($n \times 1$).
$\boldsymbol{x}$ is the vector of decision variables ($n \times 1$).
$\boldsymbol{B}$ is an ($m \times n$) constant matrix.
$\boldsymbol{h}$ is an ($m \times 1$) constant matrix.
T is the transposition of a matrix or vector.

This linear programming model has a dual form:

$$\text{Minimize} \quad \boldsymbol{h}^T \boldsymbol{\lambda} \tag{9.84}$$

$$\text{Subject to:} \quad \boldsymbol{B}^T \boldsymbol{\lambda} \ge \boldsymbol{g} \tag{9.85}$$

$$\boldsymbol{\lambda} \ge \boldsymbol{0} \tag{9.86}$$

where λis the vector of dual variables. Equation 9.85 can be replaced by the chance constraint to develop a stochastic optimization model, as follows:

$$\Pr\left[\sum_{i=1}^{m} b_{ij} \lambda_i \ge g_i \right] \ge e_j \quad \text{for } j = 1, 2, ..., n \tag{9.87}$$

which have the deterministic equivalent as follows (for more details, see Chapter 3):

$$\sum_{i=1}^{m} b_{ij} . \lambda_i \ge u_j \qquad j = 1, 2, \ldots, n \tag{9.88}$$

where e_j is the probability of exceeding ($e_j = Pr[g_j \le u_j]$), and u_j is the e_jth quantile of g_j. Therefore, the random vector $\boldsymbol{g}$ is replaced by a quantile vector $\boldsymbol{u}$ of $\boldsymbol{g}$.

Taking the dual problem of this modified LP, deterministic model proposed by Burn and McBean (1985) and Fujiwara (1988) is derived as:

$$\text{Maximize} \quad z = \boldsymbol{u}^T \boldsymbol{x} = \sum_{k=1}^{n} u_k x_k \tag{9.89}$$

$$\text{Subject to:} \quad \sum_{k=1}^{n} \sum_{i} c_k^i x_k^i \le c \tag{9.90}$$

$$x_k - \sum_{i}^{P_k} x_k^i = x_k^l \quad \forall k \tag{9.91}$$

$$0 \le x_k^i \le z_k^i \quad \forall i,k \tag{9.92}$$

$$x_k^l \le x_k \le x_k^u \quad \forall k \tag{9.93}$$

where:

u_k is the e_kth quintile of $a_k b_k$.
e_k is the probability of exceeding $a_k b_k$.
c_k^i is the ith slope of the piecewise linear treatment cost function at the kth treatment plant.
c is the budget limit.
z_k^i is the upper limit of x_k^i.
x_k^l is the lower limit of x_k.
x_k^u is the upper limit of x_k.
P_k is the number of linear segments necessary to represent the treatment cost curve for wastewater discharge k.

It is difficult to determine the probability distribution function of $a_k b_k$ when the variables such as river flow, deoxygenating and reaeration constants, and flow travel time are random. The Monte Carlo simulation can be employed to calculate the desired quantities u_k. The Monte Carlo simulation procedure to determine u_k as proposed by Fujiwara et al. (1988) is as follows:

1. Select the random variables.
2. Generate random numbers and calculate m independent values for $a_k b_k$.
3. Sort the calculated values of $a_k b_k$, (Y_1, Y_2, …, Y_m) in ascending order ($Y_1 \le Y_2 \le \ldots \le Y_m$).

4. Determine Y_i where i is the first i for which $i > me_k$. Therefore, Y_i is the e_kth quintile of $a_k b_k$.
5. For $Z_1 = Y_i$, replicate the above procedure N times to determine (Z_2, …, Z_N).
6. Calculate the mean and variance of sample Z_i:

$$\overline{Z}_N = \frac{\sum_{j=1}^{N} Z_j}{N} \tag{9.94}$$

$$S_N^2 = \frac{\sum_{j=1}^{N} \left(Z_j - \overline{Z}_N\right)^2}{(N-1)} \tag{9.95}$$

7. The relative precision (RP) is calculated as follows:

$$RP = \frac{\delta(N,\alpha)}{\overline{Z}_N} \tag{9.96}$$

where $\delta(N, \alpha)$, which is half the length of the $100(1 - \alpha)\%$ confidence interval of the expected value of Z_j, is:

$$\delta(N,\alpha) = t_{N-1,1-\frac{\alpha}{2}} \cdot \left(S_N^2 / N\right) \tag{9.97}$$

and $t_{N-1,1-(\alpha/2)}$ is the $(1 - (\alpha/2))$th quintile of the distribution with $(N - 1)$ degree of freedom.
8. If the relative precision is smaller than a given value β $(0 < \beta < 1)$, terminate the Monte Carlo simulation with $\overline{Z}_N$ being the desired quintile u_k; otherwise, increase the N to $N + 1$ to obtain the desired precision.

For more details, the reader is referred to Fujiwara et al. (1988).

9.2.6 Lake and Reservoir Water Quality Management

Water quality management in lakes and reservoirs is entirely different from rivers and streams. Reservoirs can experience dramatic changes in the physical, chemical, and biological characteristics of inflows. Phosphorus is the most significant pollutant of lakes as it can seriously affect the overall water quality. Oxygen-demanding wastes can also be important, especially when domestic or industrial wastes are discharged to a lake or upstream river. The water in reservoirs and lakes can be stratified due to a difference in density caused by temperature, dissolved substances, and suspended solids; therefore, the water quality can vary with depth, and water of varying quality can be obtained from a reservoir by withdrawing it from various levels. Water

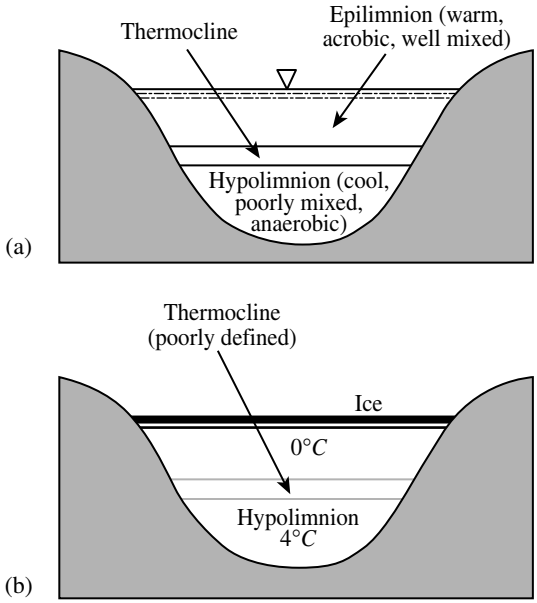


FIGURE 9.10 Stratification of a lake during (a) summer and (b) winter.

withdrawal from reservoirs can be managed in this way to supply water that satisfies water quantity and quality requirements.

9.2.6.1 Stratification and Turnover

Most lakes and reservoirs in the tropical areas are stratified during the summer and overturn in autumn, but in cold regions reservoirs can undergo winter stratification and spring overturn. Heat transfer changes the temperature at the water surface and depends on a number of factors, including air temperature, wind, humidity, and the magnitude of solar radiation. During the summer, warm water has less density and remains near the surface. The upper layer, which is nearly mixed by wind and other forces (the *epilimnion*), floats on the lower layer (the *hypolimnion*), which is cooler and poorly mixed. Figure 9.10a shows the hypolimnion and epilimnion layers in summer and winter. As shown in this figure, the boundary of these two layers is called *thermocline*, which has a high temperature and density gradient.

In the fall, as the surface layer becomes cool and winter storms arrive, the surface water that is denser than the hypolimnion sinks and the lake or reservoir experiences a fall turnover. During the turnover, the lake is isothermal from top to bottom. The turnover process stops when the temperature reaches 4°C, because water has the highest density at this condition. If the temperature of the surface water decreases, the winter stratification will occur as shown in Figure 9.10b. A spring turnover occurs when the water warms and becomes isothermal and well mixed again.

Vertical stratification affects the chemical and biological characteristics of water in the lakes and reservoirs. The thermocline restricts the diffusion of dissolved substances; therefore, the hypolimnion and the atmosphere do not contact, and

dissolved oxygen may become depleted in the hypolimnion. This condition can result in the production of hydrogen sulfide (H_2S) and the release of methane from iron-bound phosphorous of sediments, which can dramatically change the biological environment.

9.2.6.2 Eutrophication

Eutrophication is a significant factor of lakes and reservoirs that can change their ecological conditions. The introduction and cycling of nutrients can make lakes and reservoirs more productive and nutrient rich. High levels of nutrients, large populations of phytoplanktons (algae), and low transparency are characteristics of eutrophic lakes. Eutrophication time depends on the size of the lake or reservoir and the rate and characteristics of introduced nutrients. In highly eutrophic lakes, blue–green algae can result in undesirable odors and tastes and may change the appearance of the water. In deep and stratified lakes, algal production in the epilimnion can result in oxygen depletion and anaerobic conditions in the hypolimnion. Nitrogen and phosphorus are the main factors contributing to algal growth and lake eutrophication. Therefore, a major concern of water quality management in lakes and reservoirs is to limit the introduction of nutrients and to slow the entrophication rate, which is commonly accomplished by limiting phosphorus.

9.2.7 River–Reservoir Simulation Model Packages

Several computer simulation models have been developed to help water managers to develop appropriate policies to meet qualitative and quantitative objectives of the systems. These models can simulate the effects of pollution discharges into surface water bodies and evaluate temperature changes resulting from the thermal discharges or thermal stratification of reservoirs. In this section, we present the capabilities and restrictions of HEC-5Q, which has been widely used for river–reservoir quality management by different investigators.

9.2.7.1 HEC-5Q

The HEC-5Q is a simulation model for flood control and conservation systems. It has useful abilities to accept user-specified water quality and quantity objectives and to operate the network of reservoirs. The decision criteria are programmed to consider flood control, hydropower, instream flow (municipal, industrial, and irrigation water supply, and fish habitat needs), and water quality requirements (Willey et al., 1996). The first version of HEC-5Q was presented in 1979 as a modified version of HEC-5 model to evaluate reservoir operation for water quality control in large reservoir systems. The sixth phase of development of this model (1992) added the capability of simulating new qualitative parameters such as chemicals adsorbed onto organic and inorganic particles.

Deep reservoirs are conceptually modeled as a series of one-dimensional horizontal layers. Each layer is characterized by a surface area, thickness, and volume, such as those shown in Figure 9.11. Within each layer, the water is considered completely mixed, and only the vertical gradient is retained during the water quality

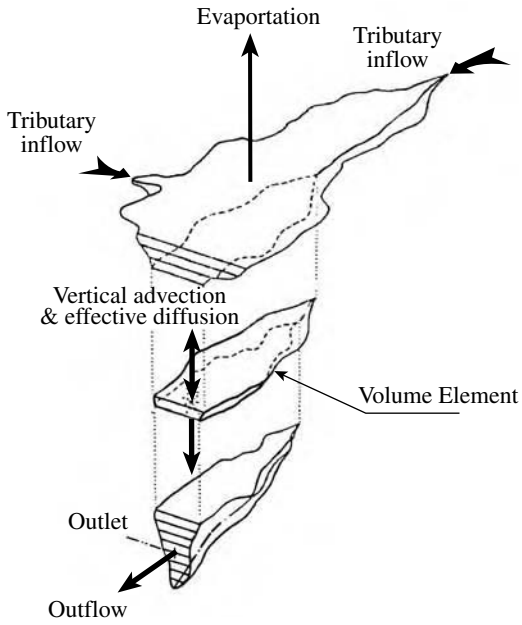


FIGURE 9.11 Geometric representation of stratified reservoir and mass transport mechanism. (From U.S. Army Corps of Engineers, *HEC5: Appendix on Water Quality Analysis*, U.S. Army Corps of Engineers, Washington, D.C., 1986.)

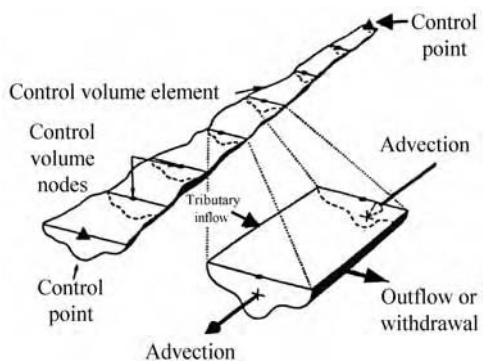


FIGURE 9.12 Geometric representation of stream system and mass transport mechanism. (From U.S. Army Corps of Engineers, *HEC5: Appendix on Water Quality Analysis*, U.S. Army Corps of Engineers, Washington, D.C., 1986.)

simulation. Withdrawals and external inflows are considered as sources or sinks of each layer.

The stream system is conceptually modeled as a linear network of volume elements. The flow simulation module requires hydraulic characteristics of a stream such as length, width, cross-section area, Manning's number, hydraulic radius, and flow and depth relationships (see Figure 9.12). The flow simulation module calculates

streamflow rates at control points using one of the several programmed hydrologic routing methods. The water quality simulation module can simulate temperature and selected conservative and nonconservative constituents that might be included in planning studies. This module computes the vertical distribution of temperature and other water quality parameters in the reservoir and downstream control points.

HEC-5Q can select the gate opening of the selective withdrawal structure to satisfy the water quality limits at downstream checkpoints. The water quality module has three modes:

1. Calibration mode
2. Annual simulation mode
3. Long-term simulation mode

In the calibration mode, decay rates and dispersion coefficients are determined based on the historical flows and water quality monitoring data. The annual simulation mode is executed on a daily basis and can evaluate the effects of reservoir operation on water quality of reservoir and downstream reaches. The time steps of the long-term simulation mode are longer (generally 30 days) and can show the effects of reservoir operation on water quality over a long-term planning horizon (e.g., more than 10 years). In the calibration and annual simulation modes, the time interval must be one day, and simulations are limited to periods contained in one calendar year (U.S. Army Corps of Engineers, 1986).

9.2.8 Reservoir Water Quality Management

The water quality in reservoirs often degrades over time because of thermal stratification and accumulation of various kinds of pollutants; consequently, reservoir releases can change the quality of downstream water bodies. The water in reservoirs usually becomes stratified due to differences in the density gradient caused by temperature or suspended and dissolved substances. The use of a selective withdrawal structure is useful for controlling the quality of releases. This structure allows the release of water from various layers of the reservoir. In this way, reservoirs can be operated to satisfy downstream water quality and quantity requirements. Numerous methods for optimal operation of reservoirs have been presented in literature, but relatively few studies have considered the qualitative and quantitative operation of reservoirs (e.g., Fontane and Labadie, 1981; Dandy and Crawly, 1992; Nandalal and Bogardi, 1994).

In the next section, a stochastic optimization model is presented for qualitative and quantitative operation of reservoirs. In this model (developed by Nandalal and Bogardi, 1994), a nonlinear optimization model at each time interval optimizes a qualitative objective function using the vertical distribution of reservoir water quality. This objective function is based on the quality of reservoir releases from outlets that are situated at different levels. Figure 9.13 shows the reservoir segments and the related outlet of each segment (layer). The volume of each layer (each smaller hypothetical reservoir) is determined based on the elevation of outlets.

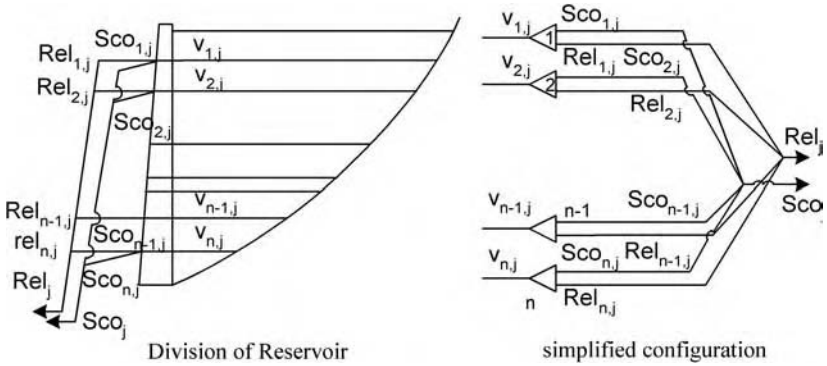


FIGURE 9.13 Division of reservoir and simplified configuration. (From Nandalal, K. D. W. and Bogardi, J. J., Reservoir management for water quality control, in *Proc. of the Ninth Congress of Asian and Pacific Division of the International Association for Hydraulic Research*, Singapore, 1994. With permission.)

It is assumed that each outlet has two openings, and the outflow of one of them should satisfy the quantitative downstream demands while the second one releases the extra water for scouring and flashing the reservoir. The following equations address the releases, scour volumes and salinity during the time period j (j = 1, 2, …, N) (Nandalal and Bogardi, 1994):

$$Rel_j = \sum_{i=1}^{n} rel_{i,j} \tag{9.98}$$

$$Sco_j = \sum_{i=1}^{n} sco_{i,j} \tag{9.99}$$

$$C_{rel,j} = \frac{\sum_{i=1}^{n} C_{i,j}.rel_{i,j}}{\sum_{i=1}^{n} rel_{i,j}} \tag{9.100}$$

$$C_{scour,j} = \frac{\sum_{i=1}^{n} C_{i,j}.sco_{i,j}}{\sum_{i=1}^{n} sco_{i,j}} \tag{9.101}$$

where:

Rel_j is the total release during period j (million m^3).
Sco_j is the total scour volume during period j (million m^3).
$C_{rel,j}$ is the release salinity during period j (mg/L).
$C_{scour,j}$ is the scour salinity during period j (mg/L).
$C_{i,j}$ is the outflow salinity during period j from outlet i (mg/L).
N is the total number of periods.
$rel_{i,j}$ is the release from outlet i during period j (million m^3).
n is the number of outlets.
$sco_{i,j}$ is the volume of scour from outlet i during period j (million m^3).

Satisfying the downstream water quality requirements and avoiding salinity buildup in the reservoir are the main qualitative objectives of reservoir operation, and the qualitative objective function can be written as:

$$\text{Minimize} \quad Z = \sum_{j=1}^{N} \left[W_1 \left(\frac{C_{rel,j} - C_{trg,j}}{C_{trg,j}} \right)^2 + W_2 \left(\frac{C_{scour,j} - \hat{C}_{trg,j}}{\hat{C}_{trg,j}} \right)^2 \right] \tag{9.102}$$

$$\text{Subject to:} \quad Rel_j = \sum_{i=1}^{n} rel_{i,j} \geq Dem_j \tag{9.103}$$

$$Rel_j + Sco_j = \sum_{i=1}^{n} \left(rel_{i,j} + sco_{i,j} \right) \leq A_j \tag{9.104}$$

$$rel_{i,j} + sco_{i,j} \leq B_i \leq V_{i,j} \qquad i = 1, 2, \ldots, n \tag{9.105}$$

$$C_{rel,j} - C_{trg,j} = 0 \qquad \text{if} \quad C_{rel,j} \leq C_{trg,j} \tag{9.106}$$

$$C_{scour,j} - \hat{C}_{trg,j} = 0 \qquad \text{if} \quad C_{scour,j} \geq \hat{C}_{trg,j} \tag{9.107}$$

where:

A_j is the allowable stochastic outflow (release + scour) during period j (million m^3), as determined from a quantitative reservoir operation optimization model.
B_i is the total allowable outflow from outlet i (million m^3).
$C_{trg,j}$ is the target release water quality (salinity) during period j (mg/L).
$\hat{C}_{trg,j}$ is the target score salinity during period j (mg/L).
Dem_j is the downstream demand of water in period j (million m^3).

$V_{i,j}$ is the initial volume of water in segment i in period j (million m^3).
W_1 and W_2 are the relative weights.

In the above nonlinear optimization model, the inflows during each time interval are neglected. The nonlinear optimization model can result in several different local optima of possible sets of releases from outlets. Some of these solutions that give the least objective function should be selected for further investigation. A reservoir water quality simulation (e.g., HEC-5Q) can be used to compare these potential sets of selective withdrawals to select the best release policy.

9.2.9 Conflict Resolution in Surface Water Quality Management

Environmental protection is commonly in conflict with other uses of water resources. In other words, environmental management is a multiple-objective problem that can be in conflict with economic, hydraulic, water quality, or environmental objectives. For example:

- Allocation of water to domestic, industrial, and agricultural demands is in conflict with instream flow and the amount of water that can be allocated to lakes and wetlands to protect their aquatic life.
- Extra discharges of water resources such as rivers and reservoirs can cause some environmental problems such as dissolved oxygen depletion.
- Agricultural return flows and industrial effluents discharged to water resources reduce the quality of water; the extent of such a reduction depends on the assimilative capacity of a particular system. Dischargers want to reduce their treatment costs, but this is not the concern of environmental agencies.

Identification of conflict issues and engaged agencies, defining their utility functions for different objectives, and defining the relative authority of agencies are the key steps for defining the required information for conflict resolution models.

The application of Nash bargaining theory (NBT) and multiple-criteria decision-making (MCDM) methods to conflict resolution of water resources has been presented in previous chapters. These methods can help to provide a better understanding of the mechanisms of cooperation and the roots of and reasons for conflict and its resolution in various political and social settings. They also can promote a better understanding of the issues, and help stakeholders more fully comprehend the potential consequences of their preferences and choices. This chapter presents another example of the application of the Nash conflict resolution method to illustrate the conflict issues and their resolution in water quality management problems.

Example 9.9

Determine the monthly allocated water to domestic, industrial, agricultural, and recreation demands in a river system shown in Figure 9.14. The river discharge is

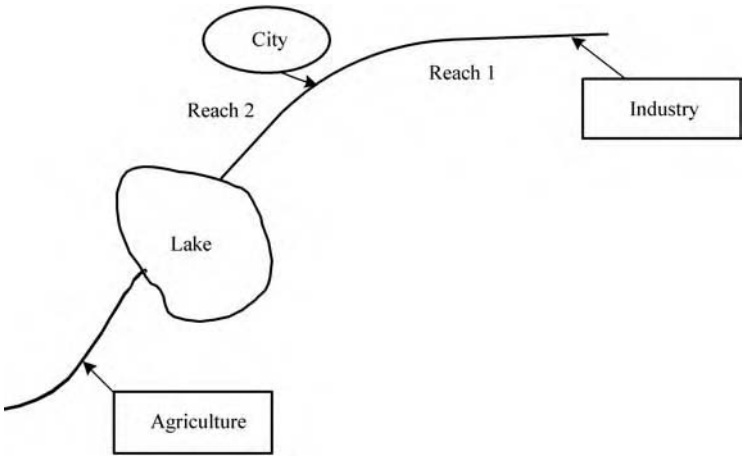


FIGURE 9.14 Components of the river system (Example 9.9).

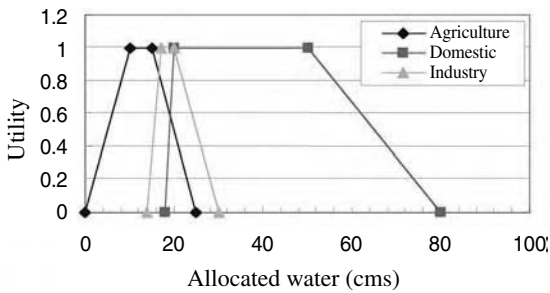


FIGURE 9.15 Utility function of water allocated to different demands.

40 m^3/sec, and the return flows of domestic and industrial sectors are 20% of the allocated water. The travel time in reaches 1 and 2 is assumed to be 8 and 4 hours, respectively. The BOD_5 concentration of headwater, domestic, and industrial return flows are 1, 30, and 50 mg/L, respectively. The BOD_5 concentration of the river at the entrance point of the lake should be less than 10 mg/L, the BOD_5 decay rate is 0.15, and the initial volume of the lake is 30 million m^3. The utility functions for agencies allocating water to different sectors are presented in Figures 9.15 and 9.16.

Solution: The nonsymmetric Nash solution of this problem is the unique optimal solution of the following problem:

$$\text{Maximize} \quad \prod_{i=1}^{5} (f_i - d_i)^{w_i}$$

$$\text{Subject to:} \quad d_i \le f_i \le f_i^* \quad \forall i$$

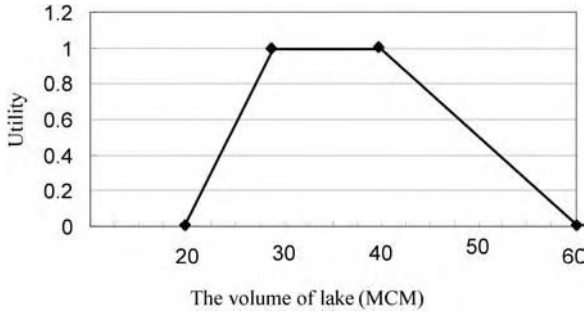


FIGURE 9.16 Utility function of the volume of a lake.

where:

w_i is the relative weight.
f_i is the utility function.
d_i is the disagreement point.
f_i^* is the ideal point of player (agency) i.

This objective function should be maximized taking into consideration the constraints of river water quality. The BOD_5 decay can be considered using Eq. (9.32). Results of the model, with and without the water quality constraint, are presented in Table 9.15. The relative weights of agencies that are related to agricultural, domestic, industrial, and recreational demands are assumed to be 0.047, 0.465, 0.023, and 0.465, respectively, and their disagreement points are 0.0, 0.9, 0.9, and 0.0, respectively. Considering that the environmental constraint and the inflow equal to 40, the BOD constraint cannot be met, and, therefore, we cannot find any conflict resolution until a hard environmental constraint is considered. Increases in upstream river discharges can greatly change the water allocated to the various sectors. Results of the conflict resolution model for river discharge of 50 m^3/sec are also presented in Table 9.15.

TABLE 9.15
Water Allocation Based on the Conflict Resolution Method

Demands	With Environmental Constraint (Inflow = 50)	Without Environmental Constraint (Inflow = 40)	Without Environmental Constraint (Inflow = 50)
Agriculture (m^3/sec)	15	8.43	14.49
Domestic (m^3/sec)	20	20	20
Industry (m^3/sec)	14	17	19.57
Final volume of lake (million m^3)	46	35	39.78

9.3 GROUNDWATER QUALITY MANAGEMENT

9.3.1 Introduction

Groundwater is a reliable source because it usually does not have high quality and quantity fluctuations. Subsurface environment is a complex system subject to contamination from numerous sources. Furthermore, the extremely slow movement of pollutants in groundwater results in a little diffusion of pollutants and a longer residence time. The restoration of groundwater quality is difficult and expensive. An understanding of the contaminants transport and degradation process is essential for groundwater quality planning and management. For this purpose, the groundwater quality simulation models that can describe how the groundwater quality system responds to, or is affected by contaminants is needed . From a planning perspective, the main issues in groundwater quality management are (Willis and Yeh, 1987):

1 To ensure that surface and subsurface waste disposal have minimal impact on the groundwater environment.
2 To optimize the waste treatment capacity of the subsurface system.

This section discusses the main groundwater pollutant sources and contaminant transport in groundwater systems. Groundwater quality management models are described in the last section of this chapter.

9.3.2 Sources of Groundwater Contamination

As water moves through the hydrologic cycle, its quality varies due to changes in physical, chemical, and biological characteristics of environment. The changes can be natural or manmade. For example, the quality of groundwater is commonly affected by waste disposal, such as storage of waste materials in excavations. Table 9.16 and Figure 9.17 show a summary of human-influenced groundwater pollutant sources. These sources can be categorized as:

1. Water-soluble products or liquid contaminants that are placed on the land or in surface water.
2. Deposited or stored substances in the ground above the water table.
3. Pollutants that are discharged in or extracted from the ground, below the water table.

For a detailed explanation of groundwater pollutants sources, the reader is referred to EPA (1985).

9.3.3 Contaminant Transport in the Saturated Zone

Groundwater quality management requires anticipating the temporal and spatial changes of contaminant concentrations in geological material. In this section, the equations of pollutant transport in surface water are extended for groundwater quality modeling. The transport and kinetic processes in groundwater are very similar to those for surface water. For groundwater quality modeling, m_s is defined as the solids

TABLE 9.16
Sources of Groundwater Quality Degradation

Groundwater Quality Problems that Originate on the Land Surface
Infiltration of polluted surface water
Land disposal of either solid or liquid wastes
Dumps
Disposal of sewage and water-treatment plant sludge
Deicing salt usage and storage
Animal feedlots
Fertilizers and pesticides
Accidental spills
Particulate matter from airborne sources
Groundwater Quality Problems that Originate in the Ground above the Water Table
Septic tanks, cesspools, and privies
Holding ponds and lagoons
Sanitary landfills
Waste disposal in excavations
Leakage from underground storage tanks
Leakage from underground pipelines
Artificial recharge
Groundwater Quality Problems that Originate in the Ground below the Water Table
Waste disposal in well excavations
Drainage wells and canals
Well disposal of wastes
Underground storage
Exploratory wells
Groundwater development

Source: EPA, *Protection of Public Water Supplies from Ground-Water Contamination*, Center for Environmental Research Information, U.S. Environmental Protection Agency, Washington, D.C.

content per unit volume of water in groundwater, and m_d is defined as the mass of solids per unit volume of space; m_s and m_d are related to each other as $\eta m_s = m_d$, where η is the soil porosity.

The Reynolds transport equation for the *dissolved pollutant concentration* (c_d) in a mass control can be written as follows (Sincero and Sincero, 1996):

For conservative substances

$$\frac{\partial(\eta c_d)}{\partial t} + \frac{\partial}{\partial x}\left(E_{xx}\frac{\partial \eta c_d}{\partial x}\right) + \frac{\partial}{\partial y}\left(E_{yy}\frac{\partial \eta c_d}{\partial y}\right) + \frac{\partial}{\partial z}\left(E_{zz}\frac{\partial \eta c_d}{\partial z}\right) + \frac{\eta}{1 + f_{cp}K_{sc}\, m_d/\eta}\left(u\frac{\partial c_d}{\partial x} + \upsilon\frac{\partial c_d}{\partial y} + \omega\frac{\partial c_d}{\partial z}\right) = 0 \tag{9.108}$$

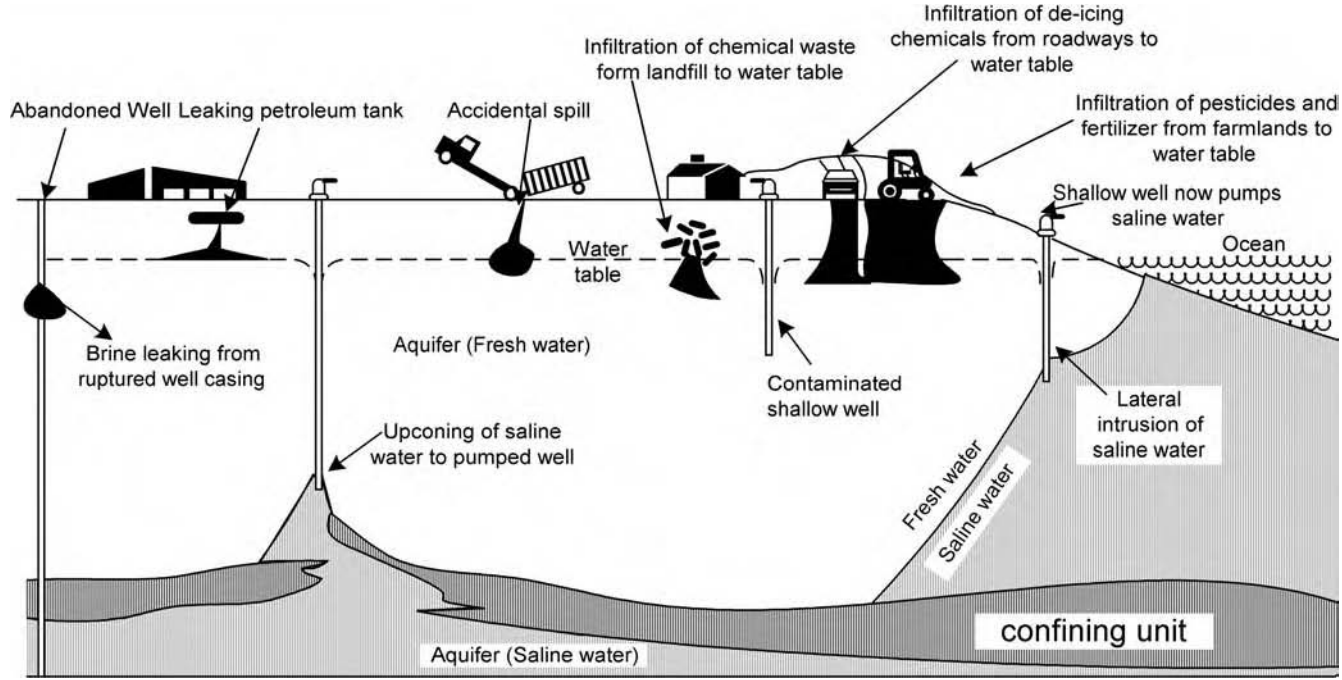


FIGURE 9.17 Sources of groundwater contamination. (From Delleur, J. W., *The Handbook of Groundwater Engineering*, CRC Press/Springer-Verlag, Boca Raton, FL, 1999. With permission.)

For nonconservative substances

$$\frac{\partial(\eta c_d)}{\partial t} - \frac{\partial}{\partial x}\left(E_{xx}\frac{\partial \eta c_d}{\partial x}\right) - \frac{\partial}{\partial y}\left(E_{yy}\frac{\partial \eta c_d}{\partial y}\right) - \frac{\partial}{\partial z}\left(E_{zz}\frac{\partial \eta c_d}{\partial z}\right) + \frac{\eta}{1 + f_{cp}K_{sc}\, m_d/\eta}\left(u\frac{\partial c_d}{\partial x} + \upsilon\frac{\partial c_d}{\partial y} + \omega\frac{\partial c_d}{\partial z}\right) + k\eta c_d + (kc_{sp}m_d = kc_d f_{cp}K_{sc}m_d) - G = 0 \tag{9.109}$$

In Eqs. (9.108) and (9.109), $kc_{sp}m_d$ represents the decay of substance sorbed onto solids. The other variables have been defined in Section 9.1.3 of this chapter.

The piezometric head h in groundwater can be calculated using Laplace equations:

$$s_c\frac{\partial h}{\partial t} + K_{xx}\frac{\partial^2 h}{\partial x^2} + K_{yy}\frac{\partial^2 h}{\partial y^2} + K_{zz}\frac{\partial^2 h}{\partial z^2} = 0 \tag{9.110}$$

where s_c is the storage coefficient. Then, u, v, and ω can be computed using Darcy's law:

$$u = \frac{K_{xx}}{\eta}\left(\frac{\partial h}{\partial x}\right) \tag{9.111}$$

$$\upsilon = \frac{K_{yy}}{\eta}\left(\frac{\partial h}{\partial y}\right) \tag{9.112}$$

$$\omega = \frac{K_{zz}}{\eta}\left(\frac{\partial h}{\partial z}\right) \tag{9.113}$$

where K_{xx}, K_{yy}, and K_{zz} are the soil hydraulic conductivities in the x, y, and z directions, respectively. Some useful equations that can be used to estimate K_{sc} are as follows (Sincero and Sincero, 1996):

$$\text{Log}\, K_{sc} = 1.00 \log K_{0,w} - 0.21 \qquad \text{(10 polyaromatic hydrocarbons)}$$
$$\text{Log}\, K_{sc} = 0.544 \log K_{0,w} + 1.377 \qquad \text{(45 organics, mostly pesticides)}$$
$$\text{Log}\, K_{sc} = 1.029 \log K_{0,w} - 0.18 \qquad \text{(13 pesticides)}$$

If an organic compound is shaked in a mixture of *n*-octanol and water, the ratio of its concentration in the octanol phase to that in the water phase is defined as $K_{0,w}$. For further discussion on nonreactive and reactive contaminant transport in saturated and unsaturated zones, the reader is referred to Delleur (1999).

9.3.4 Optimization Model for Groundwater Quality Management

Quantitative and qualitative groundwater operation is inherently a multiple-objective problem with conflicting objectives. Groundwater system optimization models can provide the optimal operational policy by considering various objectives, constraints, and their uncertainties. To satisfy the given water demand, limitation of groundwater table variation, optimal allocation of water to demands, minimization of possible contamination of the aquifer, and minimizing water or wastewater treatment costs can be the primary objectives of groundwater management models. In the next section, a typical optimization model for groundwater quality management is presented.

This model can determine the optimal waste injection concentration and pumping and injection rates considering groundwater quality standards and water demands. The objective function of the model is a weighted sum of the individual objectives:

$$\text{Maximize } Z = \sum_{t=1}^{T} \sum_{l=1}^{n} w_l . f_l\left(Q_p^t, Q_r^t, h_i^t, C_p^t, C_r^t\right) \tag{9.114}$$

where:

$f_l(Q_p^t, Q_r^t, h_i^t, C_p^t, C_r^t)$ is the lth objective.
Q_p^t is the pumping discharge from pumping site p, in planning time period t (m^3).
Q_r^t is the injection rate at injection site r in planning time period t (m^3).
w_l is the relative weight of the lth objective.
h_i^t is the groundwater head at the end of planning time period t and at control point i (m).
C_p^t and C_r^t is the pollutant concentration at pumping well/injection sites, during time period t (mg/m^3).
T is the number of time periods of the planning horizon.
n is the number of objectives.

The model constraints are as follows:

Water supply from groundwater

$$\sum_{p=1}^{P} Q_p^t \geq D^t \quad \forall t \tag{9.115}$$

$$Q_p^t \leq Q_p^{Max} \quad \forall t, p \tag{9.116}$$

where:

D^t is the total water demand in time period t (m^3).
P is the total number of pumping sites.
Q_p^{Max} is the maximum pumping rate in each time period (m^3).

Groundwater recharge

$$\sum_{r=1}^{R} Q_r^t \geq W^t \quad \forall t \tag{9.117}$$

$$Q_r^t \leq Q_r^{Max} \quad \forall t, r \tag{9.118}$$

where:

W^t is the total waste load disposal in period t (m^3).
R is the total number of injection sites.
Q_r^{Max} is the maximum injection volume in each time period (m^3).

Variations of water table

$$h_i^t \geq h_i^{Min} \quad \forall t \tag{9.119}$$

$$h_i^t \leq h_i^{Max} \quad \forall t \tag{9.120}$$

where h_i^{Min} and h_i^{Max} are the minimum and maximum head levels, respectively, at control points i (m).

Groundwater quality affecting water supply quality and resulting in clogging problems

$$C_p^t \leq C_p^{Max} \quad \forall t, p \tag{9.121}$$

$$C_r^t \leq C_r^{Max} \quad \forall t, r \tag{9.122}$$

where C_p^{Max} and C_r^{Max} are the maximum mass concentration in the pumped and injected water (mg/m^3).

The head and groundwater quality response equation can be derived from qualitative and quantitative groundwater simulation models. They can have a typical form as follows:

$$h_i^t = f_i\left(Q_p^t, Q_r^t, h_i^{t-1}\right) \qquad \forall t, i, p, r \tag{9.123}$$

$$C_i^t = g_i\left(Q_p^t, Q_r^t, C_r^t, C_i^{t-1}\right) \qquad \forall t, i, p, r \tag{9.124}$$

When nonlinear Eqs. (9.114), (9.123), and (9.124) or groundwater quality and quantity simulation models are used, the discrete dynamic programming method can be used effectively to determine operational policies and the conflict resolution for groundwater quality management. A simple example of the application of Nash's bargaining theory has been presented in Chapter 7.

9.4 PROBLEMS

9.1 If the BOD_5 of wastewater discharge by an industry is 300 mg/L and the ultimate BOD is 380 mg/L, what is the constant rate of decay? If the temperature increases to 25°C, calculate the new constant rate and ultimate BOD.

9.2 Evaluate the water quality of a river that has the following characteristics and comment on supplying agricultural and domestic demands from this river.

TDS = 364 mg / L	$Ca^{2+} = 40$ mg / L	$Mg^{+2} = 34$ mg / L	$Na^{+} = 25$ mg / L
$K^{+} = 25$ mg / L	$HCO_3^{-} = 300$ mg / L	$SO_4^{-} = 33$ mg / L	$Cl^{-} = 8$ mg / L
$NO_3^{-} = 3.3$ mg / L	PH = 7.3	EC, 25°C = 540 µs / cm	

9.3 Select seven sampling stations to monitor a river basin that has the following stream network shown here.

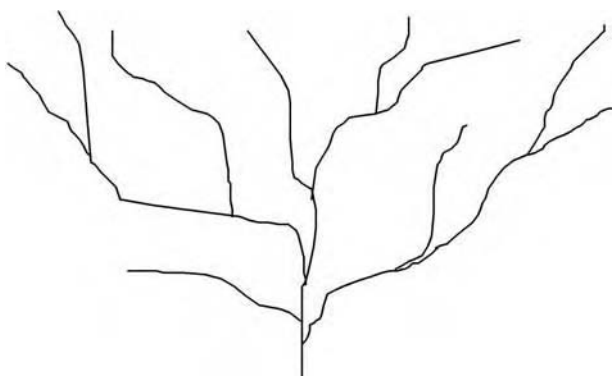


FIGURE 9.3P

9.4 Determine the mixing distance downstream of an outfall in a straight river reach that has an average width equal to 80 m and average depth equal to 6 m. The mean river velocity and the average slope of the river bed are 1 m/sec and 0.0008, respectively.

9.5 In Example 9.5, select the sampling frequency of each station using the weighting factors based on the historical means.

9.6 The statistical characteristics of total organic carbon (TOC) concentrations in five sampling stations on a river are as follows:

Station	Mean (mg/L)	Standard Deviation
1	10.0	7.8
2	11.0	3.2
3	8.0	4.5
4	9.5	2.5
5	10.1	7.0

(a) Determine the annual sampling frequencies for the stations when the width of the confidence interval of the mean at each station at the 95% confidence level is 3 mg/L.

(b) Allocate 100 samples per year to each station for a 90% confidence level.

9.7 A treatment plant plans to discharge 20,000 m^3/day partially treated wastewater into a river. Calculate the maximum BOD_5 concentration that can be discharged into the stream and the partially treated wastewater have the following characteristics:

River flow is 5 m^3/sec.
Stream temperature is 18°C.
Background BOD and DO are 2.5 and 8.4 mg/L, respectively.
Average river velocity after mixing is 0.30m/sec.
$k_c = 0.4$ and $k_2 = 2.1$ at 20°C, which is the effluent temperature.
Minimum allowable DO = 5 mg/L.

9.8 Repeat Example 9.8 considering the following items:

(a) Treatment efficiencies of Industry A and City B treatment plants can be unequal, but they have the same treatment cost.

(b) Treatment efficiencies of Industry A and City B treatment plants can be unequal, but the treatment cost in Industry A is 1.5 times the treatment cost for City B's treatment plant.

9.9 Consider a river with two point sources. Consider two reaches of lengths 20 km and 30 km and 26 checkpoints at 2-km intervals on this river (see Figure 9.9). The flow rates in a dry season have the following characteristics: The main stream flow at upstream of point source 1, q_o (m^3/sec), is assumed to be log-normally distributed and the kth wastewater flow rate, q_k (m^3/sec), is considered to be a function of q_o, as follows:

$$q_k = f_q(q_o) + \varepsilon_k$$

where ε_k is normally distributed with zero mean. The temperature is assumed to be constant and equal to 20°C. The O'Connor and Dobbins (1958) formula and Manning equation are used to calculate the reaeration constant in each reach. The cross-section of the river is considered to be

rectangular, and the river width is assumed to be large compared with flow depth. The BOD concentration of the river before the first reach, l_0, is assumed to be normally distributed. The deoxygenation constant in reach i, $k_{c,i}$, and the BOD loading of the kth discharger, b_k, are assumed to be independent and normally distributed, but the DO deficit of the kth discharger, d_k, is considered to be constant. Other information is presented in Tables 9.17 through 9.21. Calculate the optimal BOD removal rates, x_1 and x_2, for sources 1 and 2, respectively. The maximum allowable DO deficit at each control point is considered to be 3 mg/L, and the total annual budget for treatment costs is 14 units. (This problem has been adapted from Fujiwara et al., 1988.)

TABLE 9.17
Streamflow Characteristics

Variables	Unit	Mean	Standard Deviation	Lower Limit	Upper Limit
q_o	m^3/sec	45.0	20.0	20.0	—
l_o	mg/L	2.0	0.4	1.2	—
d_o	mg/L	0.0	—	—	—
ε_1	m^3/sec	0.0	2.0	–5.1	6.01
ε_2	m^3/sec	0.0	1.0	–2.5	2.50
b_1	tonnes/day	2000.0	300.0	1000.0	2500.0
b_2	tonnes/day	500.0	45.0	70.0	710.0
d_1	mg/L	4.0	—	—	—
d_2	mg/L	4.0	—	—	—
$k_{c,1}$	1/day	0.24	0.033	0.15	0.34
$k_{c,2}$	1/day	0.19	0.028	0.13	0.36

TABLE 9.18
Physical Parameters of the River

Parameters	Unit	Value
Width (reach 1)	m	65
Width (reach 2)	m	65
n_1	$m^{-\frac{1}{3}}s$	2.5×10^{-2}
n_2	$m^{-\frac{1}{3}}s$	2.5×10^{-2}
S_1	—	2.0×10^{-4}
S_2	—	2.0×10^{-4}

TABLE 9.19
Cumulative Treatment Cost Data

BOD Removal Range (%)	Monthly Cumulative BOD Removal Cost (Unit)	
	Treatment Plant 1	Treatment Plant 2
35–50	4.17	5.13
51–65	4.52	5.42
66–75	5.23	5.98
76–80	6.32	6.76
81–85	8.23	8.17
86–90	8.76	12.57
91–95	19.92	20.94
96–98	29.13	27.73

TABLE 9.20
Monte Carlo Simulation Data

Variable	Value
m	300
N_o (minimum number of iterations)	40
α	0.10
β	0.15
e_1	0.9
e_2	0.9

TABLE 9.21
Functional Relationship for Discharge Flows, Reaeration Constant, and Flow Travel Time

Variable	Value
q_1	$f_1(q_0) = 0.26\ q_0 + \varepsilon_1$
q_2	$f_2(q_0) = 0.12\ q_0 + \varepsilon_2$
k_{21}	$33.22(q_0 + q_1)^{-0.7}$
k_{22}	$34.87(q_0 + q_1 + q_2)^{-0.7}$
t_{1j}	$9.62 \times 10^{-2}\ s_{1j}(q_0 + q_1)^{-0.4}$
t_{2j}	$9.62 \times 10^{-2}\ s_{2j}(q_0 + q_1 + q_2)^{-0.4}$

Note: s_{ij}: Distance from the upstream boundary of the reach i to the checkpoint j.

TABLE 9.A1
Summary of EPA's National Primary Drinking Water Standards (Maximum Contaminant Level Goals [MCLG], Maximum Contaminant Levels [MCL])

Contaminant	MCLG (mg/l)	MCL (mg/l)
Inorganics		
Antimony	0.006	0.006
Asbestos (>10 μm)	7 MFL[a]	7 MFL[a]
Barium	2	2
Beryllium	0.004	0.004
Cadmium	0.005	0.005
Chromium (total)	0.1	0.1
Copper	1.3	TT[b]
Cyanide	0.2	0.2
Fluoride	4	4
Lead	0	TT[b]
Mercury (inorganic)	0.002	0.002
Nitrate	10	10
Nitrite	1	1
Selenium	0.05	0.05
Thallium	0.0005	0.002
Coliform and surface water treatment		
Giardia lambia	0 detected	TT[b]
Legionella	0 detected	TT[b]
Standard plate count	N/A	TT[b]
Total coliform	0 detected	—[c]
Turbidity	N/A	TT[b]
Viruses	0 detected	TT[b]
Radionuclides		
Radium 226 + radium 228	0	5 pCi/l
Gross alpha particle activity	0	15 pCi/l
Beta particle + photon radioactivity	0	4 mrem/yr

[a] MFL = million fibers per liter.

[b] TT = treatment techniques in lieu of a numerical standard.

[c] No more than 5% of the samples can be total coliform positive.

Source: Delleur, J. W., *The Handbook of Groundwater Engineering*, CRC Press/Springer-Verlag, Boca Raton, FL, 1999. With permission.

TABLE 9.A2
A Summary of Quality Tolerance For Industrial Process Waters

Industry or Use	Turbidity (mg/L)	Color (mg/L)	Hardness as $CaCO_3$ (mg/L)	Iron as Fe (mg/L)	Manganese as Mn (mg/L)	Total Solids (mg/L)	Other Requirements
Confectionery	—	—	—	0.2[a]	0.2	100	Potable
Food, general	10	—	—	0.2[a]	0.2	—	Potable
Ice	5	5	—	0.2[a]	0.2	—	Potable, $SiO_2 < 10$ mg/L
Laundering	—	—	50	0.2[a]	0.2	—	No grit or corrosiveness
Plastic, clear, uncolored	2	2	—	0.02[a]	0.02	200	No slime formation
Paper and pulp ground wood	50	20	180	1.0[a]	0.5	—	$Al_2O_3 < 8$ mg/L, $SiO_2 < 25$ mg/L, Cu <
Kraft pulp	25	15	100	0.2[a]	0.1	300	5 mg/L, pH 7.8 to 8.3
Soda and sulfite	15	10	100	0.1[a]	0.05	200	
High-grade light papers	5	5	50	0.1[a]	0.05	200	
Rayon (Viscose) pulp production	5	5	8	0.05[a]	0.03	100	

[a] Limit given applies to both iron alone and the sum of iron and manganese.

Source: Corbitt, R. A., *Standard Handbook of Environmental Engineering*, McGraw-Hill, New York, 1990. With permission.

TABLE 9.A3
Selected Guidelines for Assessing the Suitability of Water for Irrigation Purposes

Potential Problem	Variables[a]	Units[b]	Degree of Restriction of Use		
			None	Slight to Moderate	Severe
Salinity (affects availability of crop water)	EC_w	dS/m	<0.7	0.7–3.0	>0.3
	TDS	g/m^3	<450	450–2000	>2000
Permeability (affects infiltration rate of water into soil, evaluated using SAR and EC_w together)	SAR and EC_w	SAR	0–3	0–3	0–3
		dS/m	>0.7	0.7–0.2	<0.2
	SAR and EC_w	SAR	3–6	3–6	3–6
		dS/m	>1.2	1.2–0.3	<0.3
	SAR and EC_w	SAR	6–12	6–12	6–12
		dS/m	>1.9	1.9–0.5	<0.5
	SAR and EC_w	SAR	12–20	12–20	12–20
		dS/m	>2.9	2.9–1.3	<1.3
Specific ion toxicity (affects sensitive crops)	Sodium	SAR	<3	3–9	>9
	Chloride	Meq/l	<4	4–10	>10
	Boron	mg/L	<0.7	0.7–2.0	>2.0

[a] EC_w is the electrical conductivity of irrigation water, SAR is the sodium absorption ratio.
[b] dS/m = deciSiemens per meter.

Source: Tchobanoglous, G. and Schroeder, E. D., *Water Quality: Characteristics, Modeling, and Modification*, Addison–Wesley, Reading, MA, 1985. With permission.

Primary, basic Variables	Water Quantity carrier Variables	First Level Variables
	Quantity Discharge , Water Level, Small Volume in a Body of Water	
Secondary Associated Quality Variables	Quality Variables of Aggregated Effects	Second Level Variables
Examples	Temperature, PH, Turbidity, BOD, DO, Cations Anions, Conductivity, Chlorides, Radioactivity	
Aggregate-Producing Quality Variables	Quality Variables that Produce Aggregated Effects	Third Level Variables
Example 1	Radioactivity-Producing Variables: Strontium, Cesium, Tritium, etc.	
Example 2	Turbidity-Producing Variables: Suspended Matter, Colloids, Biota Groups, Dissolved Minerals, etc.	
Specific Compounds Or Species Producing Effects in Aggregation	Most Detailed Classification of Quality Variables	Fourth Level Variables
Examples	Minerals Affecting Turbidity: Iron Oxides, Mangenese Compounds, Alumina, etc.	

FIGURE 9.A1 A hierarchical ranking of water quantity and water quality random variables for the purpose of procuring information on water quality processes. (From Sanders, T. G., Ward, R. C., Loftis, J. C., Steele, T. D., Ardin, D. D., and Yevjevich, V., *Design of Networks for Monitoring Water Quality*, Water Resources Publications, Littleton, CO, 1987. With permission.)

TABLE 9.A4
Surface Water Quality Models: Basic Information

	Model	QUAL2E	SMPTOX3	HSPF	WASP5	EXAMS	CEQUALRIV1	CEQUALW2	CEQUALICM	HEC5Q	SALMONQ	MIKE11
Water body type	Stream, river	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Lake, reservoir	N	N	N	Y	Y	N	Y	Y	Y	N	N
	Estuary	Y	N	N	Y	N	Y	Y	Y	N	N	N
Dimensions	One-dimensional, branching	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Two-dimensional, X/Y	N	N	N	Y	Y	N	N	Y	N	N	N
	Two-dimensional, X/Z	N	N	N	Y	Y	N	N	Y	N	N	N
	Three-dimensional, box	N	N	N	Y	Y	N	N	Y	N	N	N
Time	Steady	Y	Y	N	N	Y	N	N	N	N	N	N
	Quasi-dynamic	Y	N	Y	Y	Y	N	N	N	N	Y	Y
	Dynamic	N	N	Y	Y	N	Y	Y	Y	Y	Y	Y
Hydrodynamics	Input	Y	Y	Y	Y	Y	N	N	N	N	N	Y
	Simulated	N	N	Y	Y	N	Y	Y	Y	Y	Y	Y
	Control structure	N	N	Y	N	N	Y	Y	Y	Y	Y	Y
Transport	Advection	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Dispersion	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
	Benthic exchange	N	Y	Y	Y	Y	N	Y	Y	N	Y	Y
Loading	Input, steady	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Input, variable	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Simulated	N	N	Y	Y	N	N	N	N	N	Y	Y
Other	Preprocessor	Y	Y	N	Y	Y	N	N	N	N	Y	Y
	Postprocessor	Y	Y	Y	Y	Y	N	N	N	N	Y	Y

Source: Mays, L. W., *Water Resources Handbook*, McGraw-Hill, New York, 1996. With permission.

TABLE 9.A5
Surface Water Quality Models: Variables and Processes

Processes	Model	QUAL2E	SMPTOX3	HSPF	WASP5	EXAMS	CEQUALRIV1	CEQUALW2	CEQUALICM	HEC5Q	MIKE11	SALMONQ
Chemical process	First-order decay	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Process kinetics	N	N	Y	Y	Y	N	N	N	N	N	N
	Daughter products	N	N	Y	Y	Y	N	N	N	N	N	N
	Sorption	N	Y	Y	Y	Y	N	N	N	N	Y	N
Sediment process	Input rates	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
	Noncohesive processes	N	N	Y	N	N	N	N	N	N	Y	N
	Cohesive processes	N	N	Y	N	N	N	N	N	N	Y	N
Water quality processes	Temperature	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y
	Salinity	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y
	Bacteria	N	N	Y	N	N	Y	Y	Y	Y	Y	Y
	DO, BOD	Y	N	Y	Y	N	Y	Y	N	Y	Y	Y
	DO, carbon balance	N	N	N	N	N	N	Y	Y	N	N	N
	Nitrogen cycle	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y
	Phosphorus cycle	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y
	Silicon cycle	N	N	N	N	N	N	Y	Y	N	Y	Y
	Phytoplankton	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y
	Zooplankton	N	N	Y	N	N	N	N	Y	N	Y	N
	Benthic algae	N	N	Y	N	N	N	N	N	N	Y	Y
	Simulate SOD	N	N	N	Y	N	N	Y	Y	N	N	N

Source: Mays, L. W., *Water Resources Handbook*, McGraw-Hill, New York, 1996. With permission.

REFERENCES

Ambrose, R. B., Wool, T. A., Connolly, J. P., and Schanz, R.W., *WASP4: A Hydrodynamic and Water Quality Model*, EPA/600/3-87/039, U.S. Environmental Protection Agency, Athens, GA, 1987.

Brown, L. C. and Barnwell, T. O., The Enhanced Stream Water Quality Models QUAL2E and QUAL2E-UNCAS Documentation and User Manual, 1987.

Burn, D. H., Water quality management through combined simulation optimization approach, *ASCE J. Environ. Eng.*, 115(5), 1011–1024, 1989.

Burn, D. H. and Lence, B. J., Comparison of optimization formulation for waste load allocation, *ASCE J. Environ. Eng.*, 118(4), 1992.

Burn, D. H. and McBean, E. A., Optimizing modeling of water quality in an uncertain environment, *Water Resources Res.*, 21(7), 934–940, 1985.

Cardwell, H. and Ellis, H., Stochastic dynamic programming models for water quality management, *Water Resources Res.*, 29(4), 803–813, 1993.

Corbitt, R. A., *Standard Handbook of Environmental Engineering*, McGraw-Hill, New York, 1990.

Dandy, G. and Crawley, P., Optimum operation of multiple reservoir system inducing salinity effects, *Water Resources Res.*, 28(4), 979–990, 1992.

Davis, M. L. and Cornwell, D. A., *Environmental Engineering*, McGraw-Hill, New York, 1991.

Delleur, J. W., *The Handbook of Groundwater Engineering*, CRC Press/Springer-Verlag, Boca Raton, FL, 1999.

Durrans, S. R., Low flow analysis with a conditional Weibull tail model, *Water Resources Res.*, 32(6), 1749–1759, 1996.

EPA, *Protection of Public Water Supplies from Ground-Water Contamination*, Center for Environmental Research Information, U.S. Environmental Protection Agency, Washington, D.C., 1985.

Fontane, D. G., Labadie, J. W., and Loftis, B., Optimal control of reservoir discharge quality through selective withdrawal, *Water Resources Res.*, 17(6), 1594–1602, 1981.

Fujiwara, O., Puangmaha W., and Hanaki, K., River basin water quality management in stochastic environment, *ASCE J. Environ. Eng.*, , 114(4), 864–877, 1988.

Houlihan, M. F. and Lucia, P. C., Groundwater monitoring, *The Handbook of Groundwater Engineering*, Edited by J. W. Delleur, CRC Press and Springer-Verlag, Chapter 24, 1999.

Hudak, P. F. and Loaiciga, H. A., An optimization method for monitoring network design in multilayered groundwater flow systems, *Water Resources Res.*, 29(8), 2835–2845, 1993.

Krenkel, P. A. and Novotny V., *Water Quality Management*, Academic Press, San Diego, CA, 1980.

Langbein, W. B. and Durum, W. H., *The Aeration Capacity of Streams*, U.S. Geological Survey Circular No. 542, U.S. Department of Interior, Washington, D.C., 1967.

Loaiciga, H. A., Charbeneau, R. J., Everett, L. G., Fogg, G. E., Hobbs, B., and Rouhani, S., Review of ground water quality monitoring network design, *ASCE J. Hydraulic Eng.*, 118(1), 11–37, 1992.

Lohani, B. N. and Thanh, N. C., Stochastic programming model for water quality management in a river, *J. Water Pollution Control Fed.*, 50, 2175–2182, 1978.

Loucks, D. P., ReVelle, C. S., and Lynn, W. R., Linear programming models for water pollution control, *Manage. Sci.*, 14(4), B166–B181, 1967.

Mays, L. W., *Water Resources Handbook*, McGraw-Hill, New York, 1996.

McKee, J. E., The need for water quality criteria, in *Proc. of a Conf. on the Physiological Aspects of Water Quality*, Division of Water Supply and Pollution Control, U.S. Public Health Service, Washington D.C., 1960.

Nandalal, K. D. W. and Bogardi, J. J., Reservoir management for water quality control, in *Proc. of the Ninth Congress of Asian and Pacific Division of the International Association for Hydraulic Research*, Singapore, 1994.

O'Connor, D. J. and Dobbins, W. E., Mechanisms of reaeration in natural streams, *Trans. Am. Soc. Civil Eng.*, 123, 655, 1958.

ReVelle, C. and McGarity, A. E., *Design and Operation of Civil and Environmental Engineering Systems*, John Wiley & Sons, New York, 1997.

Sanders, T. G., Ward, R. C., Loftis, J. C., Steele, T. D., Ardin, D. D., and Yevjevich, V., *Design of Networks for Monitoring Water Quality*, Water Resources Publications, Littleton, CO, 1987.

Sharp, W. E., A topologically optimum water-sampling plan for rivers or streams, *Water Resources Res.*, 7(6), 1641–1646, 1971.

Sincero, A. P. and Sincero, G. A., *Environmental Engineering: A Design Approach*, Prentice-Hall, Englewood Cliffs, NJ, 1996.

Streeter, H. W., and Phelps, E. B., A study of the pollution and natural purification of Ohio River, *U.S. Public Health Service Bull.*, 146, 1925.

Takyi, A. K. and Lence, B. J. (1999), Surface water quality management using multiple realization chance constraint method, *Water Resources Res.*, 35(5), 1999.

Tchobanoglous, G. and Schroeder, E. D., *Water Quality: Characteristics, Modeling, and Modification*, Addison-Wesley, Reading, MA, 1985.

U.S. Army Corps of Engineers, *HEC5: Appendix on Water Quality Analysis*, U.S. Army Corps of Engineers, Washington, D.C., 1986.

Ward, R. C., Loftis, J. C., Nielsen, K. S., and Anderson, R. D., Statistical evaluation of sampling frequencies in monitoring networks, *J. Water Pollution Control Fed.*, 51(9), 2292–2300, 1979.

Willey, R. G., Smith, D. J., and Duke, J. H., (1996), *Modeling Water Resources Systems for Water Quality Management*, *ASCE J. Water Resources Planning Manage.*, 122(3), 171–179, 1996.

Willis, R. and Yeh, W. W.-G., *Groundwater Systems Planning and Management,* Prentice-Hall, Englewood Cliffs, NJ, 1987.

10 Hydroelectric Systems Analysis

10.1 INTRODUCTION

At the United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro, Brazil, in June of 1992, various countries presented their plans for sustainable development, and most of these plans had energy sustainability as the nucleus of other activities. Definitions of *sustainable development* depend upon the subject matter, but the definition we will use here is "development by using renewable natural resources in a manner that does not eliminate or degrade them or otherwise diminish their 'renewable' usefulness for future generations…" (Karamouz, 1993).

Among all of the energy production methods, electricity, the most environmentally sound energy production method, has been given greater attention in recent years. Development of hydropower electricity generation, which has high efficiency (compared to thermal and hydrothermal units) and contributes negligible pollution to the environment, is a major attribute to the sustainable development.

Electric energy production is estimated to account for about 10% of the world's energy production (World Resources Institute, 1998), and hydropower electricity production contributes about 20% of that total (Mays, 1996). Table 10.1 shows the variability of electric energy production throughout the world. Europe, with electric energy production of 15.1%, has the highest ranking among all continents.

The published figures of electric energy production show that in developed countries, hydropower electric energy production accounts for about 3% of total energy production, whereas in oil importing and exporting developing countries, hydropower electricity production has been about 4.2 and 0.5% of total energy production, respectively. This chapter presents the concerns of water resources engineers regarding the design and operation of hydroelectric power plants and the framework of long-term and short-term operation optimization models.

10.2 BASIC DEFINITIONS

Some of the basic definitions in the area of hydroelectric systems analysis can be summarized as follows:

Gross head (H) is the difference in elevation between the water surface in at the intake and the tailwater surface.

Net head (H_n) is the head available for energy production after deducting hydraulic losses.

TABLE 10.1
Total Energy and Electric Energy Production

Region	Commercial Energy Production[a] (peta-joules[c])	Primary Electricity Production[b] (peta-joules)	Percent of Primary Electricity (%)
Africa	22,667	324	1.4
Europe	94,496	14,302	15.1
North America	89,924	11,147	12.4
Central America	8992	554	6.2
South America	17,937	1784	9.9
Asia	122,438	6957	5.7
Oceania	8436	237	2.8
World	364,891	35,305	9.7

[a] Commercial energy production includes total energy production: solid, liquid, and gases fuels and primary electricity production.
[b] Primary electricity refers to electricity generated by noncombustible energy sources and includes nuclear, wind, tidal, wave, solar, geothermal, and hydroelectric power sources.
[c] 1 peta-joule = 10 joules.

Source: Adapted from World Resources Institute (1998).

Hydraulic efficiency (e_h) is the ratio of net head to gross head:

$$e_h = H_n / H \tag{10.1}$$

Overall efficiency (e_t) is the hydraulic efficiency multiplied by the efficiencies of turbines (e_{tur}) and generators (e_g):

$$e_t = e_{tur} \times e_g \times e_h$$

Capacity is the maximum amount of power that a generator or power plant can deliver at any given time (in kW = 1000 watts).

Installed capacity is the nominal capacity of a power plant and is equal to aggregate rating of all generating equipment installed in the plant.

Generation is the energy output of a power plant.

Power is the rate of energy production, P (in kW), which can be estimated as follows:

$$P(\text{kW}) = \frac{e \cdot \gamma(\text{N}/\text{m}^3) \cdot Q(m^3/s) \cdot H_n(\text{m})}{1000} \tag{10.2}$$

where γ is the specific weight of water (9810 N/m^3) and Q is discharge through turbine (in m^3/sec) and $e = e_{tur} \times e_g$.

Energy is capable of doing work. It is exchange of the potential energy of water stored in a reservoir with the turbo generation. The energy production of a hydropower plant can be estimated using the following equation:

$$E = \int P \cdot t = \int e_t \cdot \gamma \cdot Q(t) \cdot H(t) \cdot dt \tag{10.3}$$

where E is the electrical output of the power plant (kWh) and e_t is the overall efficiency of the power generation. As can be seen in this equation, increasing head $H(t)$ and discharge $Q(t)$ through turbines increases the energy production. If the head and discharge are measured in N discrete time intervals, the energy production can be estimated as follows:

$$E = \sum_{t=1}^{N} e_t \cdot \gamma \cdot \overline{Q}(t) \cdot \overline{H}(t) \tag{10.4}$$

where $\overline{Q}(t)$ and $\overline{H}(t)$ are the average head and discharge through turbine in discrete time interval t.

Load is the demand for electricity which can be expressed in terms of energy demand (average power demand) or capacity demand (peak power demand). It can also be defined as the sequence of instantaneous power levels over a period that a power system must meet.

Load factor is the ratio of average power demand to peak power demand for a specific period that can be computed on a daily, weekly, monthly, or annual basis.

Firm or primary power is the electric energy that should be made available to a customer to meet any or all of the agreed-upon portion of that customer's load requirements. For hydropower plants, it usually can be estimated based on the energy output of the plant in the most critical period (low flow) of the historical records.

Secondary power is all power available in excess of firm power.

Plant factor is the ratio of the average load on a plant for a particular time period to its installed capacity.

Base load is the minimum load of a power system in a specific period of time.

Peak load is the maximum load in a specific period; the peaking portion of the load is usually defined as the load that occurs at the peak demand hours, which is about 8 hours a day.

Intermediate load is that portion of the load between the base load and the peaking portion of the load.

10.3 COMPONENTS OF HYDROPOWER PLANTS

Hydroelectric power plants can be classified in a number of different ways (Linsley et al., 1992):

- In terms of capacity
  - Microhydro
  - Macrohydro
- In terms of head
  - *Low head* — usually operated under heads of less than 20 m
  - *Medium head* — usually operated under heads between 20 and 30 m
  - *High head* — usually operated under heads of more than 30 m
- In terms of layout and operative mode
  - *Run-of-river* —very limited storage capacity; head varies with discharge
  - *Storage* — reservoir sufficient to provide seasonal regulation for carryover storage from the wet seasons to the dry seasons to increase energy generation in the dry season
  - *Pumped storage* — designed to convert low-value, off-peak energy to high-value, on-peak energy; water is pumped at off-peak hours from the tailwater pool to the headwater pool for energy generation in peak hours

Basic elements of a hydropower plant are as follows (see Figure 10.1) (Mays, 1996):

- The *reservoir* creates the necessary head that will provide the energy required for driving turbines.
- The *intake structure* directs water from the reservoir into the penstock or conduit. Gates or valves are used to control water discharge to the power plant. Racks or screens are also used to prevent trash or debris from entering the turbines.
- The *conduit* conveys water from the intake structure to the powerhouse.
- The *power plant* includes the turbines, generators, control, and auxiliary equipment.

Energy production is a major application of water in many river–reservoir systems. In hydropower plants, the energy in falling water runs electric turbines and generates electricity without actually consuming water. As shown in Figure 10.1,

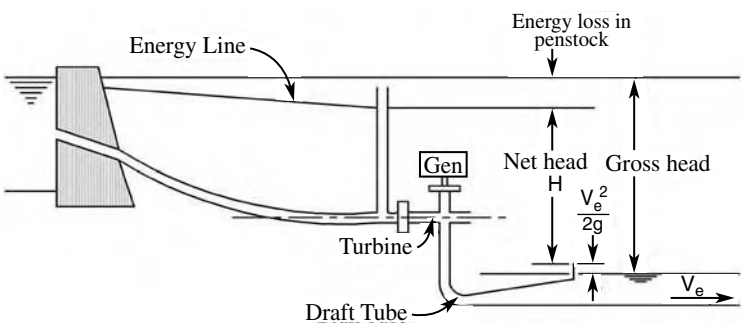


FIGURE 10.1 Element of a typical hydropower plant.

the dam provides the required head. The efficiency of power generation in the power plant shown in Figure 10.1 depends upon:

- Plant efficiency
- Volumetric water flow through the turbine
- Net hydraulic head of water across the turbine
- Type of turbine

10.4 DATA REQUIREMENTS

The following sections cover the data required to estimate the power potential of a new hydroplant or to develop operating policies for an existing hydropower plant.

10.4.1 Hydrologic Data

Historical streamflow records at the diversion point or inflows to a reservoir are the basic information required for hydropower plant development or operation studies. The proper time interval for the hydrologic data depends on factors such as the type of plant and the hydrologic characteristics of streamflows. Monthly average streamflows are often used for preliminary or even advanced studies if the streamflow does not change significantly from day to day. The degree of at-site and upstream regulation is also important in the selection of time intervals. Water losses due to evaporation from the reservoir and leakage through or around the dam and other appurtenant facilities should also be considered. In addition to water losses from the reservoir, water that is diverted for consumption or water that is reserved for operation of navigational locks or fish passage facilities should also be taken into account when estimating losses of the hydropower system.

10.4.2 Physical Characteristics

Data needed about physical characteristics of the plant facilities depends on the type of project. For storage projects, the elevation–area–volume curve and the minimum and maximum allowable reservoir elevations should be determined, whereas for run-of-river projects, elevation–discharge curves are usually helpful. Elevation–area–volume curves represent the head and volume of water stored in the reservoir that could be used for power generation. Tailwater data are also important when estimating the electricity generation of a hydropower plant. A typical tailwater rating curve is shown in Figure 10.2 which can be used for estimating tailwater elevation for different rates of plant discharge. Tailwater elevation depends on downstream channel geometry, project discharge, and downstream backwater effects. In hydropower plants with considerable head, the variation of tailwater is negligible.

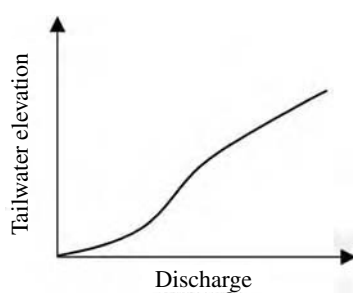


FIGURE 10.2 Typical tailwater rating curve.

Relevant data about the physical characteristics of power plants can be summarized as follows:

- Installed capacity, which represents the upper limit of power that can be generated by the plant
- Plant factor, which represents the peak and firm energy production of the plant
- Maximum and minimum discharge
- Head range that the plant operation is satisfactory
- Turbine characteristics, including shape of efficiency curve and minimum discharge

Head losses due to friction in the intake structures and penstock should also be taken into account in order to determine the net head of the system. The relation between net head and head losses can be written as follows:

$$H_n = H - H_f - V_e^2/2g \tag{10.5}$$

where H_f represents the friction losses in the trash-rack, intake structure, and penstock; $V_e^2/2g$ is the velocity at the draft tube exit.

10.4.3 Power Load and Reliability of Power Systems

Estimation of future power loads of a specific system is a primary step for hydropower development planning or operation management studies. Load predictions are usually based on the rate at which the energy consumption has increased over the past few years. Three methods have been used for estimating future power loads (American Society of Civil Engineers, 1989):

- *Trend analysis*, based on extending historical trends and modifying the projections to reflect expected changes
- *End-use analysis*, based on the expected use of electricity by different users
- *Econometric analysis,* based on the relationships between electricity demand and various factors that influence demand

In systems expansion planning, a load–resources analysis should be utilized to determine the *need* and *timing* of the output of a proposed hydropower project. Need is a measure of deficits that may occur in the future due to increased power loads compared with the current supply capacity. Timing refers to the periods when the need for additional generation occurs.

In addition to annual and monthly variation of power loads, weekly and daily load configuration are also necessary in order to determine the type of the load that a hydropower project could carry and to estimate the benefits of the system expansion

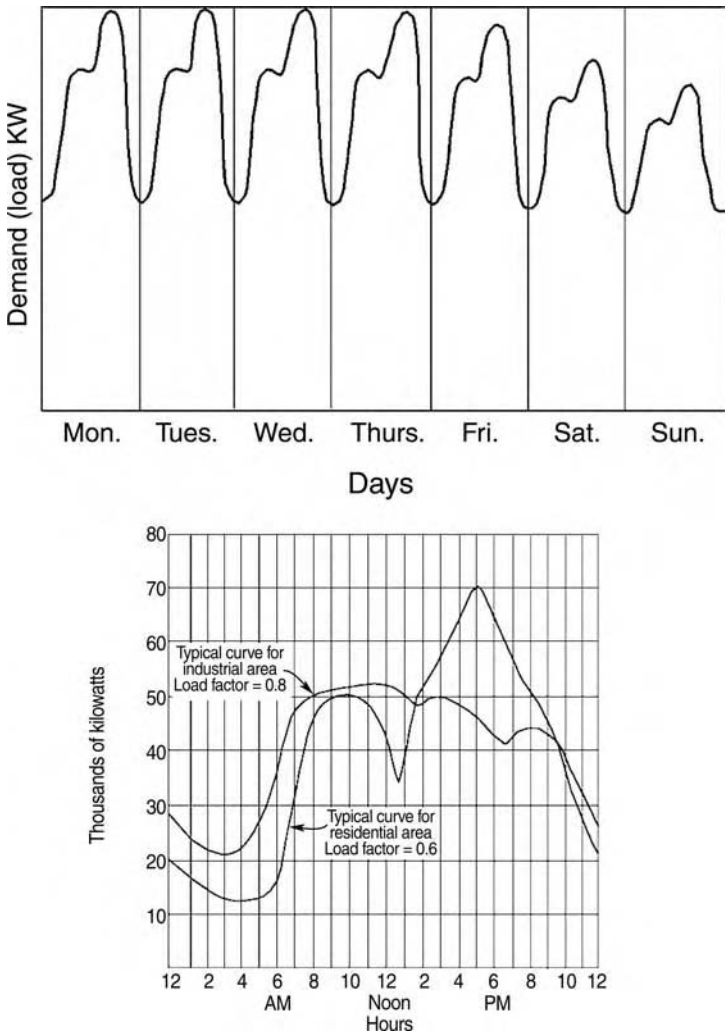
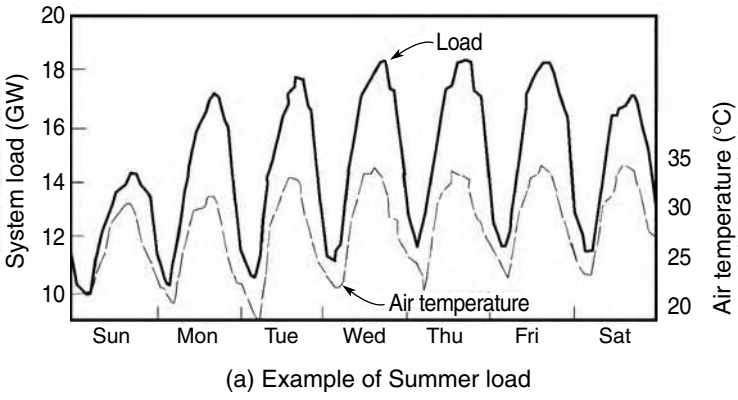


FIGURE 10.3 Typical daily and weekly load curves for an electric power system. (From Lindsley et al., 1992.)

or operation management. Figure 10.3 shows typical daily and weekly load curves. As can be seen in this figure, the demand for electricity varies from a minimum in the early hours of the morning to peak loads in the late morning or early evening. Two samples of weekly summer and winter loads for the southeastern United States are shown in Figure 10.4, where the load shape in summer has a closer correlation with air temperature which makes it easier to predict load fluctuations.

Power system operators usually divide the load into three segments (Mays, 1996):



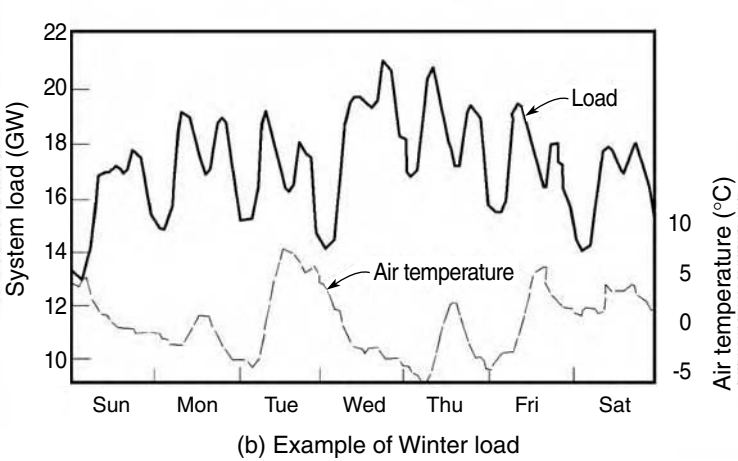


FIGURE 10.4 Sample of weekly loads for a utility located in the southeastern United States. (From Gulliver, J. S. and Arndt, R. E. A., *Hydropower Engineering Handbook*, McGraw-Hill, New York, 1991. With permission.)

- *Base load*, which is continuous, 24 hours a day
- *Peak load*, which is the highest portion of the load, occurring only for a few hours a day
- *Intermediate load*, which is the portion of the load between base and peak

Besides supplying the power loads, the system should have enough capacity to supply the expected peak load plus additional capacity to take care of breakdowns and necessary maintenance shutdowns which is usually defined in terms of *reliability.* Reliability in a hydropower system refers to the *adequacy* and *security* of a system. Adequacy relates to the existence of sufficient energy within the system to satisfy the power loads or system operational constraints. Security refers to the ability of the system to respond to disturbances within the system (Rangarajan et al., 1999).

10.4.4 Non-Power Operating Criteria

As mentioned in Chapter 6, a reservoir usually supplies water for various instream and offstream uses. Different criteria can be used in order to determine the system efficiency in satisfying other objectives such as flood control, supplying the water demands of users, and providing the required instream flow. For more details about objectives and criteria, see Chapter 8.

10.5 METHODS FOR ESTIMATING POWER POTENTIAL

Two methods have been widely used for primary estimation of power potential:

- Nonsequential or flow–duration curve
- Sequential streamflow routing (SSR)

The flow–duration curve method is the better method for all preliminary or screening studies. This method is also the best choice for high-head, run-of-river projects where head is generally fixed or even for low-head projects where head varies with discharge. For multipurpose storage projects, the SSR method is more appropriate and also can be used for examining the feasibility of including power at new water conservation or flood control projects. For peaking and pumped storage projects, hourly SSR routings are required (American Society of Civil Engineers, 1989). These methods are explained in the following sections.

10.5.1 The Flow–Duration Method

In this method, a flow–duration curve is developed based on the observed data. Figure 10.5 shows an example of a flow–duration curve. Streamflows are related to

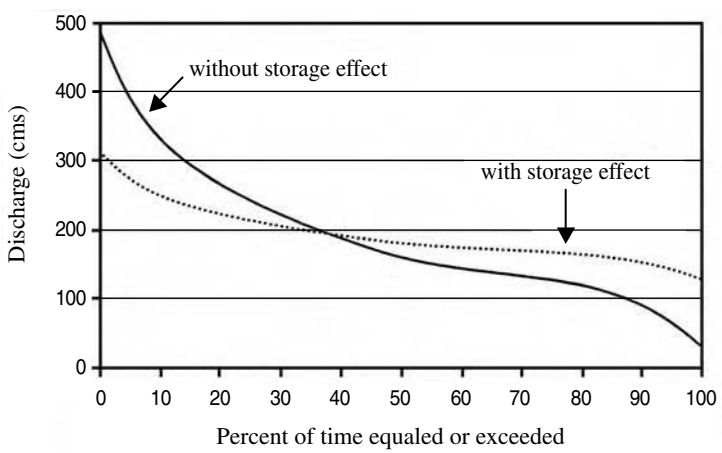


FIGURE 10.5 Typical flow–duration curve with and without storage effect.

the percentage exceedance values, which show the percent of time that different levels of streamflows are equaled or exceeded. To develop the flow–duration curves, all of the daily flow data should be ranked according to discharge, not the sequence in which they occurred. The use of daily data for development of flow–duration curves is recommended (Gulliver and Arndt, 1991). The flow–duration curve can be converted to a power–duration curve through application of the following power equation:

$$P_i(\text{kW}) = \frac{e_i \cdot \gamma \cdot Q_i \cdot H_i}{1000} \tag{10.6}$$

where:

- P_i is the power production (in kW) when the turbine discharge is at exceedance percentage i.
- e_i is the overall plant efficiency with turbine discharge equal to Q_i and net head equal to H_i.
- Q_i is the turbine discharge at percentage exceedance i (Q_{10}, Q_{20}, …); turbine discharge is assumed to be equal to river discharge except when river discharge exceeds turbine capacity or other constraints on turbine discharge are encountered.
- H_i is the net head available with river flow at exceedance percentage i.

Figure 10.6 shows the flow–duration and power–duration curves developed for St. Cloud Dam (Gulliver and Arndt, 1991). The area under the power–duration curve is the average annual energy production.

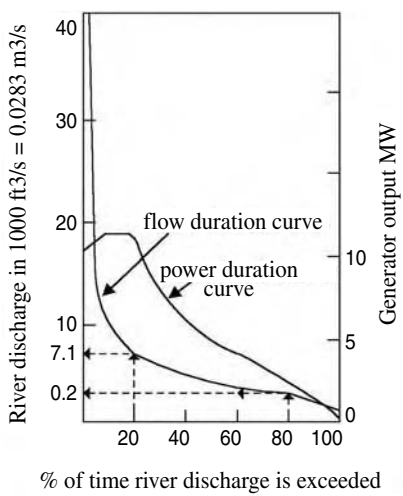


FIGURE 10.6 Flow–duration curve at St. Cloud Dam. (From Gulliver, J. S. and Arndt, R. E. A., *Hydropower Engineering Handbook*, McGraw-Hill, New York, 1991. With permission.)

TABLE 10.2
Monthly Streamflow Data for Examples 10.1 and 10.2

Month	Streamflow Data (m^3/sec) 1999	2000	2001
January	105	440	102
February	108	275	860
March	645	337	640
April	1000	515	690
May	1308	1968	1330
June	419	1965	930
July	89	360	155
August	28	71	92
September	32	50	50
October	40	95	88
November	80	100	305
December	210	88	300

Example 10.1

Table 10.2 provides the monthly streamflow data for a gauging station for the years 1999 to 2001. Plot the flow–duration curve for this river.

Solution: To develop the flow–duration curve, the observed streamflows should be arranged in descending magnitude, as shown in Table 10.3. The streamflows have been equal to or exceeded 28 m^3/sec. As can be seen in Table 10.3, the data are ranked from 1 to 36. The probability of exceedence is then estimated by the following relation:

$$Pe_i = \frac{i}{N+1}$$

where i is the rank of the data and N is the total number of data (36 in this example). The flow duration curve in Figure 10.7 shows the values in column 2 of Table 10.3 vs. the values in column 3 of this table.

Example 10.2

A run-of-river power plant is proposed at the site for which the monthly flow data are presented in Table 10.2. The head available at the site is about 10 m and plant efficiency is about 70%. Assume that the turbine discharge capacity is 500 m^3/sec. Plot the power–duration curve and find the firm energy that is expected with 90% probability of exceedance.

Solution: Using Eq. (10.6), the power–duration curve can be estimated. For example, the power output of the proposed plant for a flow of 419 m^3/sec, as in June 1999, can be determined as follows:

TABLE 10.3
Monthly Streamflow Data in Descending Order of Magnitude for Example 10.1

Rank	Sorted Streamflows	Probability of Exceedance
1	1968	0.03
2	1965	0.05
3	1330	0.08
4	1308	0.11
5	1000	0.14
6	930	0.16
7	860	0.19
8	690	0.22
9	645	0.24
10	640	0.27
11	515	0.30
12	440	0.32
13	419	0.35
14	360	0.38
15	337	0.41
16	305	0.43
17	300	0.46
18	275	0.49
19	210	0.51
20	155	0.54
21	108	0.57
22	105	0.59
23	102	0.62
24	100	0.65
25	95	0.68
26	92	0.70
27	89	0.73
28	88	0.76
29	88	0.78
30	80	0.81
31	71	0.84
32	50	0.86
33	50	0.89
34	40	0.92
35	32	0.95
36	28	0.97

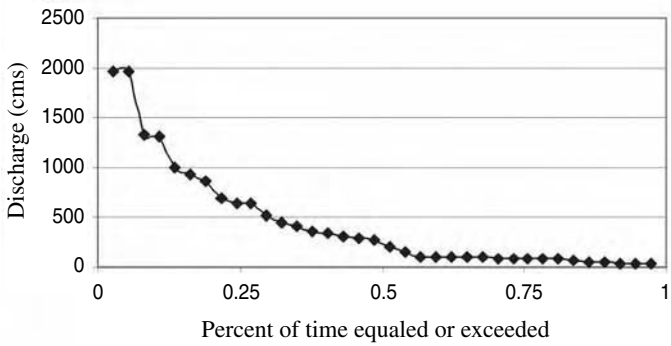


FIGURE 10.7 Flow–duration curve for Example 10.2.

$$P = e \cdot \gamma \cdot Q \cdot H = 0.7 \times 9.81 \times 419 \times 10 = 28,773 \text{ kW}$$

The power generation in each of the months is estimated using the above relation and shown in Table 10.4. It should be noted that in the months when the discharge is higher than the maximum turbine discharge (500 m^3/sec), the maximum power generation can be estimated as:

$$P_{\max} = e \cdot \gamma \cdot Q \cdot H = 0.7 \times 9.81 \times 500 \times 10 = 34,335 \text{ kW}$$

Figure 10.8 shows the resulting power–duration curve. The minimum power generation has been 1922.8 kW. The firm energy with 90% reliability is then estimated as 2967 kW using the numbers in Table 10.4.

Example 10.3

Assume that we are sizing the turbines in the proposed power plant for Example 10.2 to run full 30% of the time. Find the turbine design discharge.

Solution: The turbine design discharge can be found by taking the intercept of the 30% exceedance ordinate on the flow–duration curve and moving horizontally across to the river discharge abscissa at 515 m^3/sec. This is the river discharge that is met or exceeded 30% of the time and will be the turbine design discharge.

10.5.2 Sequential Streamflow Routing (SSR) Method

Sequential streamflow routing, which was primarily developed for evaluating storage projects, is based on the continuity equation:

$$\Delta S = I - O - L \tag{10.7}$$

TABLE 10.4
Monthly Power Generation (kW) Based on the Streamflow Data in Table 10.3 (Example 10.1)

Rank	Sorted Streamflows	Probablity of Exceedance	Power Generation (kW)
1	1968	0.03	34335
2	1965	0.05	34335
3	1330	0.08	34335
4	1308	0.11	34335
5	1000	0.14	34335
6	930	0.16	34335
7	860	0.19	34335
8	690	0.22	34335
9	645	0.24	34335
10	640	0.27	34335
11	515	0.30	34335
12	440	0.32	30215
13	419	0.35	28773
14	360	0.38	24721
15	337	0.41	23142
16	305	0.43	20944
17	300	0.46	20601
18	275	0.49	18884
19	210	0.51	14421
20	155	0.54	10644
21	108	0.57	7416
22	105	0.59	7210
23	102	0.62	7004
24	100	0.65	6867
25	95	0.68	6524
26	92	0.70	6318
27	89	0.73	6112
28	88	0.76	6043
29	88	0.78	6043
30	80	0.81	5494
31	71	0.84	4876
32	50	0.86	3434
33	50	0.89	**3434**
34	40	0.92	**2747**
35	32	0.95	2197
36	28	0.97	1923

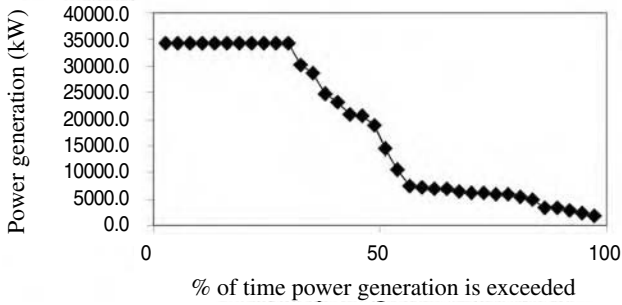


FIGURE 10.8 Power duration curve for Example 10.2

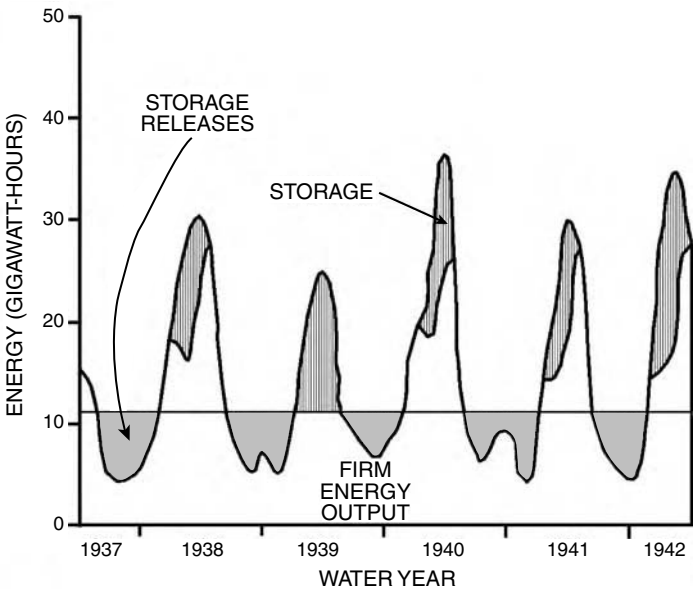


FIGURE 10.9 Effect of storage on firm energy production of a hydropower plant. (From American Society of Civil Engineers, *Civil Engineering Guidelines for Planning and Designing Hydroelectric Developments*, Vol. 1, *Planning, Design of Dams, and Related Topics, and Environmental*, American Society of Civil Engineers, New York, 1989. With permission.)

where I and O are inflow to and outflow from the reservoir; L represents the losses due to evaporation, seepage, etc.; and ΔS is the change in reservoir storage. This equation is applied sequentially to all historic records in order to obtain a continuous record of project operation. Hourly, daily, weekly, or monthly data can be used depending on the nature of the study and the type of data available (American Society of Civil Engineers, 1989). Energy output can then be estimated by applying the reservoir outflow values to the power equation, Eq. (10.2).

The SSR method can be used for maximizing firm energy production. In run-of-river plants, the firm energy that can be produced is equal to the energy production

in the most critical low-flow periods in the historical records. However, in storage projects water can be stored in the high-flow periods in order to increase plant discharge in low-flow periods. Figure 10.9 shows the effect of storage on firm energy production of a typical storage power plant.

Example 10.4

Assume a reservoir with 8 billion m^3 of storage is constructed at the site discussed in Example 10.2. The firm yield of the reservoir is considered to be 400 m^3/sec, which is also equal to the discharge capacity of the turbine. Assume that the total hydraulic losses are negligible. The storage-water elevation relation at the reservoir site is given as follows:

$$h = 4 \times (S + 7)$$

where h is the water elevation in the reservoir (in meters) and S is the reservoir storage (in billion cubic meters). Use the following relation for estimating the tailwater elevation:

$$h_T = \left(R + R^s\right) \times 0.0009$$

where R and R^s are the regulated and spilled releases (in m^3/sec), respectively, from the reservoir, and h_T is the tailwater head (in meters). Consider that the reservoir is full at the beginning of the first month. Use the SSR method to estimate the expected monthly power generation in the reservoir.

Solution: Table 10.5 shows the results of the SSR method. For example, examine the data for the first month. The reservoir storage was assumed to be at full capacity, or 8 billion m^3. When taking into consideration the 400-m^3/sec constant release and 105-m^3/sec inflow to the reservoir, the reservoir storage at the end of this month can be estimated as:

$$\text{Storage at the end of Jan.} = 8 + (105 - 400) \times \frac{31\left(\frac{\text{days}}{\text{mon}}\right) \times 24\left(\frac{\text{hrs}}{\text{day}}\right) \times 60\left(\frac{\text{mins}}{\text{hr}}\right) \times 60\left(\frac{\text{secs}}{\text{min}}\right)}{10^9}$$

$$= 7.21 \text{ Billion cubic meters}$$

The elevation at the beginning and the end of January can be found as:

$$\text{Elevation at the beginning of Jan.} = 4 \times (8 + 7) = 60 \text{ m}$$

$$\text{Elevation at the end of Jan.} = 4 \times (7.21 + 7) = 56.84 \text{ m}$$

TABLE 10.5
Results of the SSR Method in Example 10.4

Month	Storage (10^9 m^3)	Begin. Elev. (m)	Inflow (cms)	Inflow (MCM)	Release (controlled) (cms)	Release (controlled) (MCM)	Release (Spilled) (cms)	Release (Spilled) (MCM)	Ending Stor. (10^9 m^3)	Ending Elev. (m)	Tail-water Elev. (m)	Net Head (m)	Energy (MWh)
Jan.	8.00	60.00	105	281.23	400	1071.36			7.21	56.84	0.36	58.06	114,824.52
Feb.	7.21	56.84	108	270.60	400	1002.24			6.48	53.91	0.36	55.02	108,805.35
Mar.	6.48	53.91	645	1727.57	400	1071.36			7.13	56.54	0.36	54.87	108,507.01
Apr.	7.13	56.54	1000	2505.60	400	1002.24	637.80	0.25	8.00	60.00	0.36	57.91	114,525.73
May	8.00	60.00	1308	3503.35	400	1071.36	2431.99	908.00	8.00	60.00	1.18	58.82	116,333.62
Jun.	8.00	60.00	419	1086.05	400	1036.80	49.25	19.00	8.00	60.00	0.38	59.62	117,915.97
Jul.	8.00	60.00	89	238.38	400	1071.36			7.17	56.67	0.36	57.97	114,655.02
Aug.	7.17	56.67	28	75.00	400	1071.36			6.17	52.68	0.36	54.32	107,419.23
Sep.	6.17	52.68	32	82.94	400	1036.80			5.22	48.87	0.36	50.41	99,705.34
Oct.	5.22	48.87	40	107.14	400	1071.36			4.25	45.01	0.36	46.58	92,118.59
Nov.	4.25	45.01	80	207.36	400	1036.80			3.42	41.69	0.36	42.99	85,023.94
Dec.	3.42	41.69	210	562.46	400	1071.36			2.91	39.66	0.36	40.31	79,730.30
Jan.	2.91	39.66	440	1178.50	400	1071.36			3.02	40.09	0.36	39.51	78,141.18
Feb.	3.02	40.09	275	689.04	400	1002.24			2.71	38.83	0.36	39.10	77,326.12
Mar.	2.71	38.83	337	902.62	400	1071.36			2.54	38.16	0.36	38.14	75,419.86
Apr.	2.54	38.16	515	1290.38	400	1002.24			2.83	39.31	0.36	38.37	75,892.15
May	2.83	39.31	1968	5271.09	400	1071.36			7.03	56.11	0.36	47.35	93,643.46

(continued)

TABLE 10.5 (CONTINUED)
Results of the SSR Method in Example 10.4

Month	Storage (10^9 m^3)	Begin. Elev. (m)	Inflow (cms)	Inflow (MCM)	Release (controlled) (cms)	Release (controlled) (MCM)	Release (Spilled) (cms)	Release (Spilled) (MCM)	Ending Stor. (10^9 m^3)	Ending Elev. (m)	Tail-water Elev. (m)	Net Head (m)	Energy (MWh)
Jun.	7.03	56.11	1965	5093.28	400	1036.80	3083.79	1189.73	8.00	60.00	1.43	56.62	111,984.78
Jul.	8.00	60.00	360	964.22	400	1071.36			7.89	59.57	0.36	59.43	117,526.02
Aug.	7.89	59.57	71	190.17	400	1071.36			7.01	56.05	0.36	57.45	113,616.79
Sep.	7.01	56.05	50	129.60	400	1036.80			6.10	52.42	0.36	53.87	106,543.00
Oct.	6.10	52.42	95	254.45	400	1071.36			5.29	49.15	0.36	50.42	99,723.46
Nov.	5.29	49.15	100	259.20	400	1036.80			4.51	46.04	0.36	47.24	93,416.54
Dec.	4.51	46.04	88	235.70	400	1071.36			3.67	42.70	0.36	44.01	87,035.46
Jan.	3.67	42.70	102	273.20	400	1071.36			2.88	39.50	0.36	40.74	80,573.04
Feb.	2.88	39.50	860	2154.82	400	1002.24			4.03	44.11	0.36	41.45	81,974.89
Mar.	4.03	44.11	640	1714.18	400	1071.36			4.67	46.69	0.36	45.04	89,076.36
Apr.	4.67	46.69	690	1728.86	400	1002.24			5.40	49.59	0.36	47.78	94,493.04
May	5.40	49.59	1330	3562.27	400	1071.36			7.89	59.56	0.36	54.21	107,219.65
Jun.	7.89	59.56	930	2410.56	400	1036.80	1262.82	487.20	8.00	60.00	0.80	58.98	116,643.93
Jul.	8.00	60.00	155	415.15	400	1071.36			7.34	57.38	0.36	58.33	115,354.23
Aug.	7.34	57.38	92	246.41	400	1071.36			6.52	54.08	0.36	55.37	109,495.68
Sep.	6.52	54.08	50	129.60	400	1036.80			5.61	50.45	0.36	51.90	102,644.36
Oct.	5.61	50.45	88	235.70	400	1071.36			4.78	47.10	0.36	48.42	95,750.66
Nov.	4.78	47.10	305	790.56	400	1036.80			4.53	46.12	0.36	46.25	91,471.32
Dec.	4.53	46.12	300	803.52	400	1071.36			4.26	45.05	0.36	45.22	89,437.93

TABLE 10.6
Rule Curve Policy for Reservoir Operation for Example 10.5

	Month											
	Jan.	**Feb.**	**Mar.**	**Apr.**	**May**	**June**	**July**	**Aug.**	**Sept.**	**Oct.**	**Nov.**	**Dec.**
S_t [a]	5	6	7	8	8	7	6	5	6	6	6	6

[a] S_t is the reservoir storage at the beginning of month t (10^9 m^3).

The average water elevation in the reservoir is (60 + 56.84)/2 = 58.42 m. The tailwater elevation in this month can be found as:

$$h_T = (400 + 0) \times 0.0009 = 0.36 \text{ m}$$

Therefore, the net head in this month would be 58.42 – 0.36 = 58.06 m. The power generation in this month can then be estimated as:

$$E = \frac{e \cdot \gamma \cdot Q \cdot H}{1000} \times t = \frac{0.7 \times 9.81 \times 400 \times 58.06}{1000} \times 720 = 114,825 \text{ MWh}$$

Example 10.5

In Example 10.4, assume that the reservoir should be operated based on the rule curve shown in Table 10.6. Assume also that the maximum turbine discharge is 700 m^3/sec. Determine the monthly power generation of the reservoir.

Solution: Table 10.7 shows the results of the SSR method. For example, examine the first month. The reservoir storage was 5 billion m^3 at the beginning of the month. Based on the rule curve policy, the reservoir storage at the end of January should be 6 billion m^3. The inflow volume in this month is 281 million m^3, and the storage at the end of this month is estimated to be 5000 + 281 = 5281 million m^3. Therefore, the release is set to 0 for this month. The estimated monthly power generation values for the rest of the operation period are shown in Table 10.7.

10.6 HYDROPOWER RESERVOIR OPERATION

Firm energy generation in most runoff river hydropower plants is very low due to seasonal variations of streamflows (low discharge in dry seasons) and low head. Storage power plants provide more energy with a relatively high head and less variable discharge. In general, through storage and pondage in hydropower projects the following objectives can be attained:

- Increasing firm energy
- Maximizing average annual energy
- Maximizing dependable capacity

TABLE 10.7
Results of the SSR Method in Example 10.5

Month	Storage (10^9 m^3)	Begin. Elev. (m)	Inflow (cms)	Inflow (MCM)	Total Release (MCM)	Release (Controlled) (cms)	Release (Controlled) (MCM)	Release (Spilled) (cms)	Release (Spilled) (MCM)	Ending Stor. (10^9 m^3)	Ending Elev. (m)	Tail-water Elev. (m)	Net Head (m)	Energy (MWh)
Jan.	5.00	48.00	105	281.23	0.00	0.00	0.00	0.00	0.00	5.28	49.12	0.00	49.12	0
Feb.	5.28	49.12	108	270.60	0.00	0.00	0.00	0.00	0.00	5.55	50.21	0.00	50.21	0
Mar.	5.55	50.21	645	1727.57	0.00	0.00	0.00	0.00	0.00	7.28	57.12	0.00	57.12	0
Apr.	7.28	57.12	1000	2505.60	1785.00	688.66	1785.00	0.00	0.00	8.00	60.00	0.62	59.38	202,184
May	8.00	60.00	1308	3503.35	4503.35	700.00	1874.88	981.36	2628.47	7.00	56.00	1.51	54.49	188,577
Jun.	7.00	56.00	419	1086.05	2086.05	700.00	1814.40	104.80	271.65	6.00	52.00	0.72	51.28	177,463
Jul.	6.00	52.00	89	238.38	1238.38	462.51	1238.38	0.00	0.00	5.00	48.00	0.42	47.58	108,813
Aug.	5.00	48.00	28	75.00	0.00	0.00	0.00	0.00	0.00	5.07	48.30	0.00	48.30	0
Sep.	5.07	48.30	32	82.94	0.00	0.00	0.00	0.00	0.00	5.16	48.63	0.00	48.63	0
Oct.	5.16	48.63	40	107.14	0.00	0.00	0.00	0.00	0.00	5.27	49.06	0.00	49.06	0
Nov.	5.27	49.06	80	207.36	0.00	0.00	0.00	0.00	0.00	5.47	49.89	0.00	49.89	0
Dec.	5.47	49.89	210	562.46	1034.90	0.00	0.00	386.39	1034.90	5.00	48.00	0.35	47.65	0
Jan.	5.00	48.00	440	1178.50	178.50	0.00	0.00	66.64	178.50	6.00	52.00	0.06	51.94	0
Feb.	6.00	52.00	275	689.04	0.00	0.00	0.00	0.00	0.00	6.69	54.76	0.00	54.76	0
Mar.	6.69	54.76	337	902.62	0.00	0.00	0.00	0.00	0.00	7.59	58.37	0.00	58.37	0
Apr.	7.59	58.37	515	1290.38	882.04	0.00	0.00	340.30	882.04	8.00	60.00	0.31	59.69	0

May	8.00	60.00	1968	5271.09	6271.09	700.00	1874.88	1641.36	4396.21	7.00	56.00	2.11	53.89	186,521
Jun.	7.00	56.00	1965	5093.28	6093.28	700.00	1814.40	1650.80	4278.88	6.00	52.00	2.12	49.88	172,648
Jul.	6.00	52.00	360	964.22	1964.22	700.00	1874.88	33.36	89.34	5.00	48.00	0.66	47.34	163,842
Aug.	5.00	48.00	71	190.17	0.00	0.00	0.00	0.00	0.00	5.19	48.76	0.00	48.76	0
Sep.	5.19	48.76	50	129.60	0.00	0.00	0.00	0.00	0.00	5.32	49.28	0.00	49.28	0
Oct.	5.32	49.28	95	254.45	0.00	0.00	0.00	0.00	0.00	5.57	50.30	0.00	50.30	0
Nov.	5.57	50.30	100	259.20	0.00	0.00	0.00	0.00	0.00	5.83	51.33	0.00	51.33	0
Dec.	5.83	51.33	88	235.70	1069.11	0.00	0.00	399.16	1069.11	5.00	48.00	0.36	47.64	0
Jan.	5.00	48.00	102	273.20	0.00	0.00	0.00	0.00	0.00	5.27	49.09	0.00	49.09	0
Feb.	5.27	49.09	860	2154.82	428.01	0.00	0.00	176.94	428.01	7.00	56.00	0.16	55.84	0
Mar.	7.00	56.00	640	1714.18	714.18	0.00	0.00	266.64	714.18	8.00	60.00	0.24	59.76	0
Apr.	8.00	60.00	690	1728.86	1728.86	0.00	0.00	667.00	1728.86	8.00	60.00	0.60	59.40	0
May	8.00	60.00	1330	3562.27	4562.27	700.00	1874.88	1003.36	2687.39	7.00	56.00	1.53	54.47	188,508
Jun.	7.00	56.00	930	2410.56	3410.56	700.00	1814.40	615.80	1596.16	6.00	52.00	1.18	50.82	175,872
Jul.	6.00	52.00	155	415.15	1415.15	0.00	0.00	528.36	1415.15	5.00	48.00	0.48	47.52	0
Aug.	5.00	48.00	92	246.41	0.00	0.00	0.00	0.00	0.00	5.25	48.99	0.00	48.99	0
Sep.	5.25	48.99	50	129.60	0.00	0.00	0.00	0.00	0.00	5.38	49.50	0.00	49.50	0
Oct.	5.38	49.50	88	235.70	0.00	0.00	0.00	0.00	0.00	5.61	50.45	0.00	50.45	0
Nov.	5.61	50.45	305	790.56	402.27	0.00	0.00	155.20	402.27	6.00	52.00	0.14	51.86	0
Dec.	6.00	52.00	300	803.52	1803.52	0.00	0.00	673.36	1803.52	5.00	48.00	0.61	47.39	0

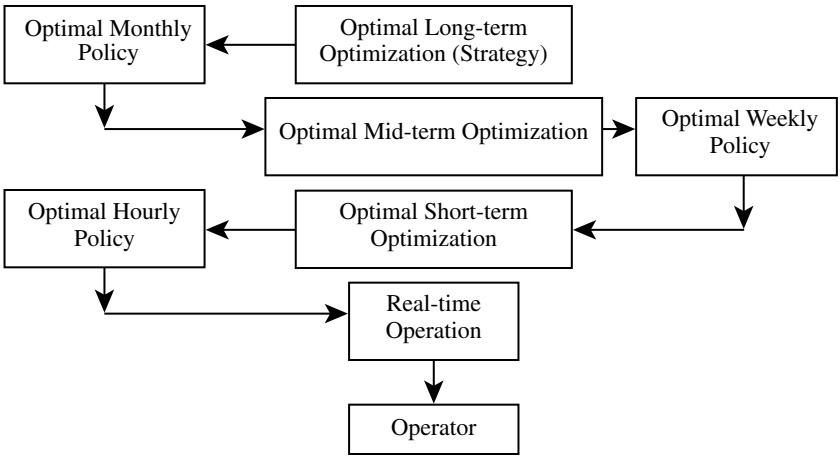


FIGURE 10.10 Flowchart comparing long-, mid-, and short-term operation optimization models for a hydropower reservoir.

Firm energy can be increased in storage power plants because the streamflows stored in high-flow season can be discharged in dry seasons at a rate higher than the normal flow rates in those seasons (see Figure 10.9). To optimize hydropower reservoir operations, maximizing the firm and/or average annual energy can be considered as the objective function. Hydropower storage projects are usually multipurpose (see Chapter 8 for discussion on various objectives of reservoir operation). Considering the variety of these objectives and the temporal variation of power loads, a time decomposition approach is necessary for hydropower reservoirs (see Figure 10.10). In the flowchart shown in Figure 10.10, the higher level policies impose constraints on the lower time sequence models (Yeh et al., 1992). As seen in this figure, optimization of the hydropower reservoirs operation consists of four steps:

1. *Long-term planning* (strategic): optimization of reservoir operation on a monthly scale within the planning time horizon
2. *Mid-term planning* (strategic and tactical): optimization of reservoir operation on a weekly scale within the 1-year time horizon
3. *Short-term planning* (tactical): optimization of reservoir operation on an hourly scale within the 1-week time horizon
4. *Real-time operation*: optimization of load dispatching and unit commitment scheduling on an hourly scale

Hydropower units often operate as a part of a larger system. Even though performance of a reservoir in supplying water demands might affect only local users downstream of the reservoir, the power generation of a hydropower plant might have more far-reaching effects on the power network of the region. In the following sections, the bases of the long-term and short-term planning and real-time operation models for multipurpose hydropower storage projects are discussed.

10.6.1 Long-Term Planning Models

The main objective of long-term planning of hydropower storage projects is to maximize the net benefits of systems operation. Reservoirs usually should satisfy such objectives as:

- Supply of different water demands
- Flood control
- Supply of power loads

With respect to supplying instream flows, a conflict arises between the water supply objectives and maintaining enough head in the reservoirs for efficient power generation. A general formulation of a long-term planning model can be summarized as follows:

$$\text{Minimize} \quad Z = \sum_{t=1}^{n} \left(loss_t^1 \left(R_t, D_t^W \right) + loss_t^2 \left(P_t, D_t^L \right) \right) \qquad (t = 1, \ldots, n) \qquad (10.8)$$

$$\text{Subject to:} \quad S_{t+1} = S_t + I_t - R_t - L_t \qquad (t = 1, \ldots, n) \qquad (10.9)$$

$$P_t = f\left(R_t^p, E_t \right) \qquad (t = 1, \ldots, n) \qquad (10.10)$$

$$E_t = g\left(S_t \right) \qquad (t = 1, \ldots, n) \qquad (10.11)$$

$$R_t = R_t^p + R_t^s \qquad (t = 1, \ldots, n) \qquad (10.12)$$

$$S_{\min} \le S_t \le Cap \qquad (t = 1, \ldots, n) \qquad (10.13)$$

$$0 \le R_t \le R_{\max,t} \qquad (t = 1, \ldots, n) \qquad (10.14)$$

$$S_t, E_t, L_t, R_t, P_t, R_t^p, R_t^s \ge 0 \qquad (t = 1, \ldots, n) \qquad (10.15)$$

where:

I_t is the reservoir inflow in month t.
R_t is the total reservoir release in month t.
S_t is the reservoir storage at the beginning of month t.
D_t^W is the water demands in month t.
D_t^L is the power load in month t.
E_t is the average net head level in month t, which is a function of reservoir storage and release.
L_t is the reservoir losses in month t due to evaporation and seepage.

Cap is the total reservoir storage.
P_t is the power generation in month t.
R_t^p is the reservoir release used for energy production in month t.
R_t^s is the reservoir release from non-power outlets in month t.
$loss_t^1$ and $loss_t^2$ are the operation losses associated with performance of the reservoir in supplying water demands and generation of power, respectively.
n is the number of months in the planning horizon.

As seen in Eq. (10.8), the objective function of the monthly optimization model consists of two parts. The first part represents operation losses based on the water supply objectives. The second part of the objective function represents losses associated with power generation and is usually estimated based on a comparison between the generated power and power load of the system or a target power generation defined by system operators and decision makers. Transmission losses should also be taken into account. A suitable way of considering transmission losses in preliminary studies is to incorporate them into a plant efficiency factor (Yeh et al., 1992).

Example 10.6

For the first year of the data given in Example 10.4, consider that a reservoir supplies the domestic demands of a city located downstream of the reservoir. It also supplies part of the energy demands of the city. Formulate an optimization model to determine the optimal releases for only the first year of data. Use the following loss functions to estimate the losses associated with shortages in supplying water demands and energy loads:

$$loss_t^1 = \begin{cases} 1.56 \times \left(D_t^W - R_t\right)^3 & \text{if} \quad D_t^W > R_t \\ 0 & \text{otherwise} \end{cases}$$

$$loss_t^2 = \begin{cases} 7.35 \times \left(D_t^L - P_t\right) & \text{if} \quad D_t^L > P_t \\ 0 & \text{otherwise} \end{cases}$$

where $loss_t^1$ and $loss_t^2$ are losses associated with water supply and energy generation objectives, respectively. D_t^W and D_t^L are water and energy demands (expressed in m^3/sec and MWh, respectively). R_t and P_t are water release and power generation, respectively, in month t.

Solution: The objective function is to minimize the losses:

$$\text{Minimize} \quad \sum_{t=1}^{12} \left[loss_t^1 + loss_t^2\right]$$

Subject to:

$$\left.\begin{aligned} S_2 &= S_1 + 281.23 - R_1 \\ S_3 &= S_2 + 270.6 - R_2 \\ S_4 &= S_3 + 1727.57 - R_3 \\ S_5 &= S_4 + 2505.60 - R_4 \\ S_6 &= S_5 + 3503.35 - R_5 \\ S_7 &= S_6 + 1086.05 - R_6 \\ S_8 &= S_7 + 238.38 - R_7 \end{aligned}\right\} \text{Constraints for monthly reservoir water balance}$$

$$\begin{aligned} S_9 &= S_8 + 75 - R_8 \\ S_{10} &= S_9 + 82.94 - R_9 \\ S_{11} &= S_{10} + 107.14 - R_{10} \\ S_{12} &= S_{11} + 207.36 - R_{11} \\ S_{13} &= S_{12} + 562.44 - R_{12} \\ S_1 &= 8000 \\ S_i &\le 8000 \quad \forall i \\ S_i, R_i &\ge 0 \end{aligned}$$

10.6.2 Mid-Term Planning Models

A mid-term planning model has a structure similar to the long-term model. The only difference between these models is in constraints and forecasts accuracy. Because the planning horizon consists of a smaller number of periods, the model is capable of further decompositions and closer simulation. It can be on a weekly or daily scale within a monthly or weekly time horizon, respectively. The long-term operation policy that defines the monthly release from reservoir and power generation should be considered as a constraint for a mid-term model.

10.6.3 Short-Term Planning Models

The purpose of the short-term optimization model is to determine the optimal operating policies on an hourly time scale within a weekly or daily time horizon. The hourly model is considerably more detailed than the long-term and mid-term planning models. Different types of generating units in each power plant and their respective input–output performance relationships should be considered. The economic objective functions could be taken into account in the short-term optimization models by incorporating the value of peak and offpeak generation. Figure 10.11 shows a schematic diagram of the value of power generation within a month. In the first two classes ($k = 1, 2$), covering about 220 hours each month, the power generated would have a high value (peak generation). The third class ($k = 3$) would have a

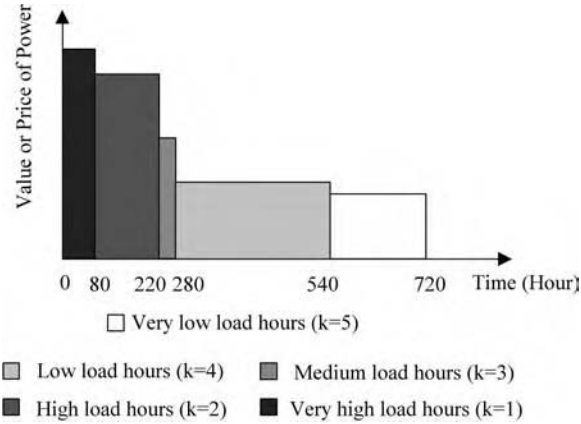


FIGURE 10.11 Schematic classification of benefits of power generation on an hourly time scale for a month. (Karamouz, M., Zahraie, B., and Araghi-Nejhad, Sh., *ASCE J. Computing Civil Eng.*, 2003 [submitted].)

medium value, and the last two classes ($k = 4, 5$) would have a very low value (firm generation). Therefore, the main objective of the short-term optimization model could be defined to maximize the peak power generation of the plant considering the optimal policies obtained from the mid-term optimization model as a constraint. The objective function of this model can be defined as:

$$\text{Maximize} \quad \sum_{k=1}^{5} \sum_{t=1}^{720} P_{k,t} \times V_k \tag{10.16}$$

where V_k is the value of power and $P_{k,t}$ is the power generation of the power plant in class k.

10.6.4 Real-Time Operation Models

The real-time operation models for hydropower units optimize the unit commitment problem, which is a complex decision-making process involving the integrated hourly scheduling of generators in a multiple-objective hydropower system. The purpose is to satisfy power loads, water demands, reliability constraints, operational restrictions, and security requirements on an hourly scale and within each day. Constraints of the unit commitment problem can be formulated as follows:

- Unit generation ($P_{t,i}$) should be less than the unit capacity:

$$P_{t,i} \le P_i^{cap} \qquad (t = 1, \ldots, 24) \tag{10.17}$$

  where $P_{t,i}$ is the power generated by unit i in time interval t and P_i^{cap} is the power generation capacity of unit i.

- Total daily releases of the units of a hydropower plant should be equal to releases determined by the short-term optimization model:

$$R_t^{os} = \sum_{t=1}^{24} \sum_{i=1}^{NU} R_{t,i} \tag{10.18}$$

where R_t^{os} is the optimal daily release obtained from the short-term optimization model, NU is the number of units in the hydropower plant, and $R_{t,i}$ is the hourly release from unit i in time t.

- System security requirements should be satisfied:

$$\sum_{i=1}^{NU} P_{t,i} \le P_t^{\max} \qquad (t = 1, \ldots, 24) \tag{10.19}$$

where $P_t^{\max}$ is the maximum allowable rate of power generation in time t.

- Release from each unit should be less than the maximum discharge capacity of the generators:

$$R_{t,i} \le R_i^{\max} \qquad (t = 1, \ldots, 24) \tag{10.20}$$

where $R_i^{\max}$ is the maximum discharge capacity of unit i.

The objective function of the unit commitment problem in a system of multiple power plants can be defined in two different ways:

- Maximize the operation efficiency of the system, as follows (Yi et al., 1997):

$$E_{total} = \frac{\sum_{t=1}^{24} \sum_{i=1}^{NU} P_{t,i}}{\sum_{t=1}^{24} \sum_{i=1}^{NU} R_{t,i} E_{t,i}} \tag{10.21}$$

where $E_{t,i}$ is the net head on the unit i in time t.

- Minimize the total cost of operation of the system considering costs of operation of each of the units.

10.7 HYDROTHERMAL COORDINATION

Optimal coordination of hydropower systems offer significant benefits, including reductions in fossil fuel consumption and the investment required for a possible expansion of the power plants (Yeh et al., 1992; Yeh and Becker, 1983). The

hydrothermal coordination problem can be classified into two categories (Wood and Wollenberg, 1996):

- When the hydroelectric component is by far the largest of the hydrothermal system, scheduling of these systems can be optimized by minimizing the costs of the thermal plants.
- Systems with a closer balance between the hydro and thermal plants and those for which the hydroelectric system is a small fraction of the total capacity can be optimally scheduled by minimizing the thermal generation production costs and considering all of the hydraulic constraints that may exist.

The hydrothermal optimization problem is of considerable difficulty because of the following factors:

- High dimensionality of the problem with regard to:
  - Seasonal state variables for reservoir operation, including storage and historical and forecasted inflows
  - Hourly state variables for short-term optimization of power plant operation
- Nonlinearity of the problems, including:
  - Power transmission losses
  - Generator input–output relationships
  - Cost functions of hydro and thermal units
- Stochastic nature of the problem variables, including:
  - Reservoir inflows
  - Power loads
- Large number of constraints for considering:
  - Supply and demand balance
  - Flow balance or continuity equation
  - Operational limits on hydro and thermal generation, including thermal plant loading and unloading limitations
  - Bounds on water release
  - Equipment maintenance schedule
  - Spinning reserve requirements

Because of these factors, a time decomposition approach is necessary. The weekly and daily cyclic nature of power loads has led researchers to specify operational periods in terms of weeks, days, and hours. The following provides details regarding short-term optimization of the hydrothermal power generation problem.

The primary objective of hydrothermal scheduling is to determine the amount of hydro and thermal generation required to meet the power loads such that the costs of operation are minimized. Operating costs of hydroplants are usually invariant and negligible compared to the costs of thermal power plants; therefore, minimization of thermal production costs, which are primarily a function of the fuel costs required

to run the thermal generators, can be considered as the objective function of this problem.

For this purpose, thermal plant cost functions, also called lambda (λ) curves, can be used. Figure 10.12 shows the typical lambda curve for the southeastern United States (Georgakakos et al., 1997). This curve shows that at higher system loads, the production cost of an additional unit of energy is higher. Therefore, hydropower plants should be scheduled to replace the more expensive thermal generation with hydrogeneration. Figure 10.13 shows the typical weekly system generation requirement and hydrothermal coordination for supplying power loads. For example, for an hourly interval k, the lambda curve can be approximated by:

$$\lambda = aL + b \tag{10.22}$$

where a and b are constant coefficients and L is the power load. As shown in Figure 10.14, the cost savings can be estimated as follows:

$$CS_k = \frac{\left[aL_k + b\right] + a\left[L_k - P_k\right] + b}{2} P_k = \frac{-a}{2} P_k^{\,2} + \left[aL_k + b\right]P_k \tag{10.23}$$

where P_k is the displaced power generation by hydropower units and CS_k is the cost savings for hourly time interval k. This approximation is based on assuming the λ curve is linear within specific intervals.

Specific constraints for optimization of hydrothermal systems should be considered. Among the different constraints for operation of thermal units, constraints for the maximum allowable change in power generation of these systems within short time intervals (e.g., hourly intervals) are the most common:

$$\begin{aligned} T_k - T_{k-1} &\le \Delta T \\ T_{k-1} - T_k &\le \delta T \end{aligned} \tag{10.24}$$

$$T_k \le T_{\max} \tag{10.25}$$

where ΔT and δT are the maximum allowable increase and decrease, respectively, in the thermal unit output at hourly time intervals; T_k is the power generation of a thermal unit at hour k; and T_{max} is the maximum capacity of the thermal unit. If $\lambda_k(T_k)$ represents the costs of a thermal unit operating at power level T_k in time interval k, the objective function of the hydrothermal coordination problem in a scheduled horizon including M intervals can be written as:

$$\text{Minimize} \quad C = \sum_{k=1}^{M} n_k \lambda_k\left(T_k\right) \tag{10.26}$$

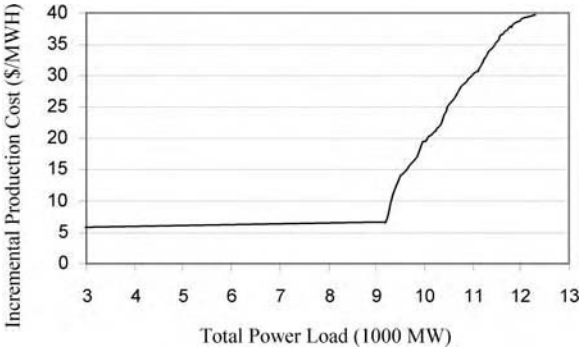


FIGURE 10.12 Typical lambda curve. (From Georgakakos, A. P. et al., *ASCE J. Water Resources Planning Manage.*, 123(1), 30–38, 1997. With permission.)

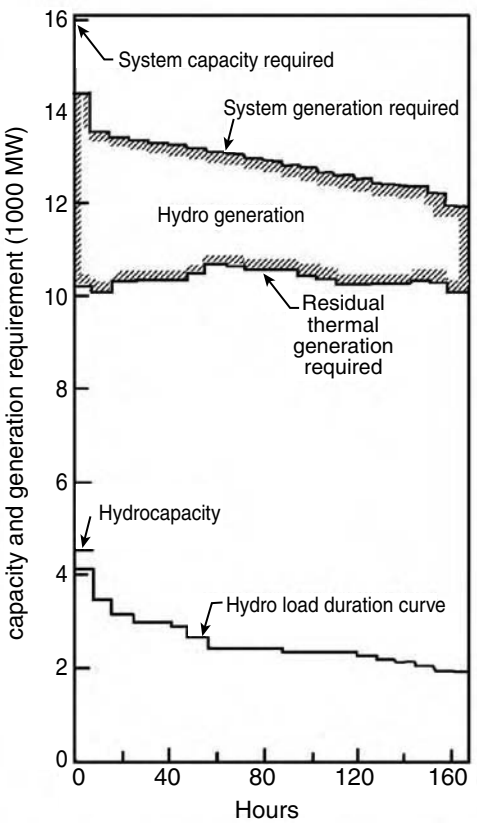


FIGURE 10.13 Typical weekly system generation requirement and hydrothermal coordination. (From Gulliver, J. S. and Arndt, R. E. A., *Hydropower Engineering Handbook*, McGraw-Hill, New York, 1991. With permission.)

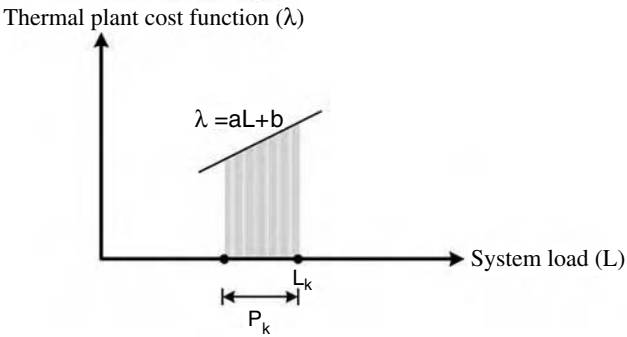


FIGURE 10.14 Approximation of thermal cost savings. (From Georgakakos, A. P. et al., *ASCE J. Water Resources Planning Manage.*, 123(1), 30–38, 1997. With permission.)

where n_k is the length of interval k in the schedule horizon. The total power load in the kth interval is:

$$L_k = P_k + T_k - P_k^l \tag{10.27}$$

where P_k^l is the electric loss between the hydroplant and the load and can be estimated as a function of hydropower generation:

$$P_k^l = h\left(P_k\right) \tag{10.28}$$

By combining Eqs. (10.27) and (10.28), we can write:

$$h\left(P_k\right) - P_k + \left(L_k - T_k\right) = 0 \tag{10.29}$$

Assuming that the hydrogeneration is a function of water discharge rate, the water discharge rate in interval k can be estimated as a function of $(L_k - T_k)$:

$$Q_K = \sum_{k=1}^{M} n_k g\left(L_k - T_k\right) \tag{10.30}$$

where Q_K is the total water discharge in schedule horizon M (Wong and Wong, 1994). It should be mentioned that in the above formulation, the transmission losses are estimated only for hydropower generation. In more sophisticated formulations, the transmission losses are considered for both hydro and thermal units (see Yeh et al., 1992 for more details).

Example 10.7

Suppose you want to optimize the energy-scheduling problem in a system for which the hydroelectric component is the primary source. Use the following notation to formulate the hydrothermal coordination problem:

L_m is the load of the system in period m ($m = 1, \ldots, M$).
T_m is the power generation by thermal units in period m.
P_m is the power generation by hydro units in period m.
T_{tot} is the total power generated by thermal units.
$\lambda(T_m)$ is the cost of the thermal units, which is a function of power generation.
C_{tot} is the total cost in the planning horizon.
M is the total number of time periods in the planning horizon ($m = 1, \ldots, M$).
i_m is the number of hours in each period.
M_T is the total number of periods for which the thermal plant is running.

Solution: It is desired to supply the entire amount of load from the hydroplant in such a way that the cost of the thermal plant is minimized. The objective function is to minimize the total cost within the planning horizon, which can be written as:

$$\text{Minimize} \quad C_{tot} = \sum_{m=1}^{M_T} \lambda\left(T_m\right) \cdot i_m$$

$$\text{Subject to:} \quad \sum_{m=1}^{M_T} T_m i_m - \sum_{m=1}^{M} L_m i_m + \sum_{m=1}^{M} P_m i_m = 0 \tag{10.31}$$

This constraint shows that the total power generation in the thermal and hydro plants should supply the total load of the system. It should be noted that all physical constraints related to power generation capacity of hydro and thermal plants should also be incorporated and are not presented here.

Example 10.8

In Example 10.7, the cost of power generation in thermal plants can be estimated using the following relation:

$$\lambda\left(T_m\right) = aT_m^2 + bT_m + c \tag{10.32}$$

Use the Lagrange function to find the optimal rate of power generation in the thermal units, assuming that the optimal rate of power generation in the thermal plant (T^*) is constant for the entire period it is operating (Wood and Wollenberg, 1996).

Solution: The Lagrange function can be written as:

$$L = \sum_{m=1}^{M} \lambda(T_m) \cdot i_m + \alpha\left(-\sum_{m=1}^{M_T} T_m i_m + \sum_{m=1}^{M} L_m i_m - \sum_{m=1}^{M} P_m i_m \right)$$

Then,

$$\frac{\partial L}{\partial T_m} = \frac{d\lambda(T_m)}{dT_m} - \alpha = 0 \qquad \text{for } m = 1, \ldots, M$$

or,

$$\frac{d\lambda(T_m)}{dT_m} = \alpha \qquad \text{for } m = 1, \ldots, M$$

This means that the thermal plant should be run at a constant incremental cost for the entire period it is on. Considering the constant rate of power generation in thermal plants, it can be written as:

$$C_{tot} = \sum_{m=1}^{M_T} \lambda(T^*) \cdot i_m = \lambda(T^*) \sum_{m=1}^{M_T} i_m = \lambda(T^*) \cdot M_T$$

Considering the cost function for the thermal plant (Eq. (10.32)), it can be written that:

$$C_{tot} = \left(aT^{*2} + bT^* + c \right) \cdot M_T$$

Using Eq. (10.31), it can be written:

$$T^* M_T - \sum_{m=1}^{M} L_m i_m + \sum_{m=1}^{M} P_m i_m = 0$$

Therefore,

$$C_{tot} = \left(aT^{*2} + bT^* + c \right) \cdot \frac{\sum_{m=1}^{M} L_m i_m - \sum_{m=1}^{M} P_m i_m}{T^*}$$

To minimize the total cost, it can be written that:

$$\frac{dC_{tot}}{dT^*} = \left(a - \frac{c}{T^{*2}}\right) \cdot \left(\sum_{m=1}^{M} L_m i_m - \sum_{m=1}^{M} P_m i_m\right) = 0$$

Therefore,

$$T^* = \sqrt{c / a}$$

The optimal rate of power generation of the thermal plant is estimated to be $\sqrt{c / a}$.

10.8 CONFLICT ISSUES IN THE OPERATION OF HYDROPOWER SYSTEMS

Hydropower plants are usually an important element of multipurpose water resources systems. The major conflict issues in operation of hydropower systems arise when a reservoir that supplies water for different activities must also provide enough discharge and head for energy supply. Monthly variations of power loads (energy demand) and water demands usually do not follow the same pattern; therefore, supplying one of these demands in a specific period of time might conflict with supplying the other demands in the future. As mentioned in Chapter 1, various organizations are responsible for water and energy supply, and the requirements and constraints for these sectors and agencies should be incorporated when formulating the conflict resolution problem. The major conflict issues in operation of hydropower reservoirs can be summarized as follows:

- Keeping enough head in reservoirs for power generation with high efficiency is in conflict with
  - Supplying water demands, especially in dry seasons
  - Keeping flood control storage
  - Improving the water quality in the reservoir
- Water and energy demands do not follow the same pattern; therefore, optimal scheduling of water release from the reservoir for these purposes might vary.

The first step for formulating the conflict resolution problem should be recognizing the main conflict issues for all the relevant organizations. For example, consider a river–reservoir system that supplies the following demands:

- Domestic water demand
- Agricultural water demand
- Industrial water demand
- Part of energy demand of a city located downstream of the reservoir

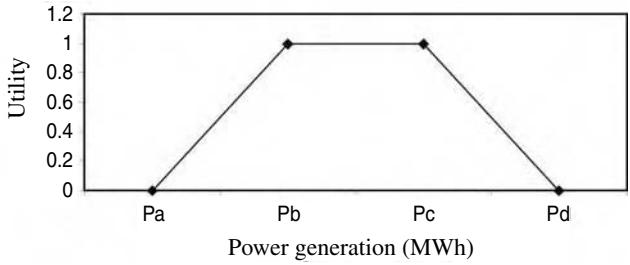


FIGURE 10.15 Utility function for a range of power generation.

Management of water supply and consumption in each of these sectors is usually done by a specific agency. For example, even though a specific department (for our purposes, the Department of Water Supply) is responsible for supplying water for a variety of demands, the Department of Agriculture is responsible for managing irrigation demands and maintaining an acceptable level of production efficiency. The Department of Industries is focused on supplying enough water for industrial production, and supplying energy demands is the main focus for the Department of Energy, whose actions could have local, regional, or national impacts. Importing energy from electricity distribution networks of other states or countries and exporting energy to these networks are other important issues for the operation of power systems. Finally, coordination of hydrothermal units is a critical component of conflict resolution in these systems.

In the second step, the priorities and favorable ranges of water and energy supply for each of these agencies should be recognized. Each of the above-mentioned departments has their own set of priorities for allocating water to different demands. For example, the Department of Energy could define the utility function for a range of energy supply, as shown in Figure 10.15. In this figure, the most favorable annual volume of water allocation for this organization ranges from P_b to P_c MWh. This range can be estimated based on the variability of energy demands and coordination with other sources of power supply such as thermal units.

The increasing segment in the left-hand side of this figure shows the range of least favorable to most favorable rate of power generation. The points on this line are not totally rejected by this organization because the shortage can be supplied from other sources, such as more expensive thermal generation. The decreasing limb on the right-hand side of this graph also shows that even though the power is generated in the hydropower plant these conditions are also less favorable because of probable problems in coordination with other sources of electricity supply and limitation in the power transmission network.

All other agencies selected in the first step should also provide a favorable range of water supply for their different demands. Besides water demands and the priorities of agencies in supplying those demands, each agency has a specific level of authority in regard to changing the water allocation schemes and imposing a favorable range of water supply within the existing political and institutional climate of a region. For example, the Department of Agriculture has an interest in supplying a relatively higher volume of water to the agricultural lands than to the industrial or municipal

sectors. Analysts can also define a set of relative weights based on the authority of the various agencies to make decisions regarding water allocation.

The final step is to formulate the conflict resolution problem. For this purpose, different methods can be used, as briefly explained in Chapter 2. The steps that should be taken for conflict resolution in hydropower systems modeling can be summarized as follows:

- The conflict issues should be recognized based on the objectives of the hydropower system operation.
- The agencies engaged in water and energy resources allocation and consumption should be recognized.
- These agencies should provide their own sets of favorable levels of water supply for their demands.
- Analysts should define the relative authority of agencies in imposing the water and energy allocated to each demand.

Example 10.9

Consider a reservoir that supplies water to the following demand points:

- A city located downstream of the reservoir
- Industries located downstream of the reservoir near the city
- Agricultural networks downstream of the reservoir near the city

A run-of-river power plant is also located on the river downstream of the water diversion for the above demand points. The head available at the site of this power plant is 10 m and plant efficiency is about 60%. The following agencies are affected by decisions made to release water from the reservoir:

- Agency 1: Department of Water Supply
- Agency 2: Department of Agriculture
- Agency 3: Department of Industries
- Agency 4: Department of Energy

The main objective is to meet the water and energy demands of these agencies. The Department of Water Supply has a twofold role — namely, to allocate water to different purposes as well as supply water for domestic purposes (*DD*). The irrigation demand (*RD*), industrial demand (*ID*), and energy demand (*ED*) should also be accounted for. Decision makers in these agencies are asked to set their most favorable ranges of water allocation and energy production by assigning either a 1 (most favorable) or a 0 (least favorable); see Figures 10.16 to 10.19. Analysts set the following weights on the relative importance of water demands within the political climate of that region:

Agricultural:	0.19
Domestic:	0.38
Industrial:	0.19
Energy:	0.24

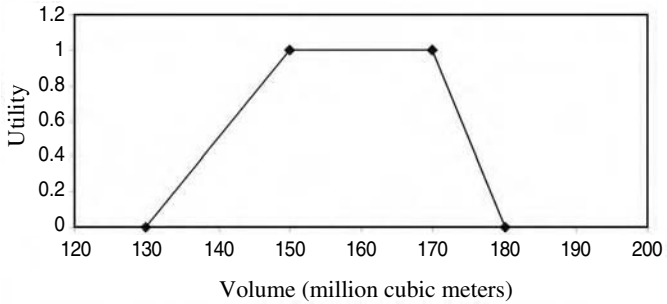


FIGURE 10.16 Utility function for water allocation for the domestic demand defined by Agency 1.

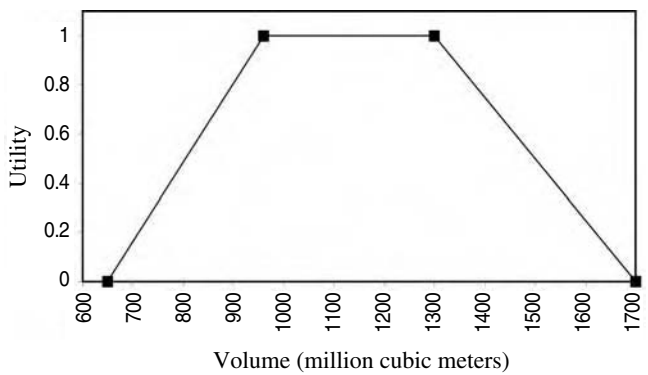


FIGURE 10.17 Utility function for water allocation for the agricultural demand defined by Agency 2.

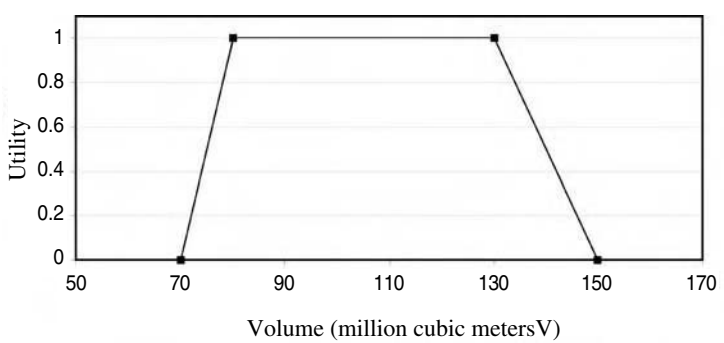


FIGURE 10.18 Utility function for water allocation for the industrial demand defined by Agency 3.

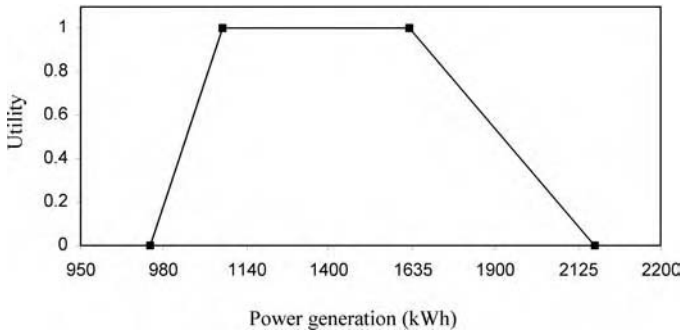


FIGURE 10.19 Utility function for the power generation defined by Agency 4.

Average annual inflow to the reservoir is estimated to be 1400 million m^3. Find the most appropriate water allocation scheme for a normal year and a year with the projected inflow of 1000 million m^3.

Solution: The discharge required for generating the range of power defined by Agency 4 (see Figure 10.19) can be estimated as follows:

$$\text{Utility} = \begin{cases} 0 & Q = 980{,}000 \Big/ (0.6 \times 9.81 \times 10 \times 24 \times 365) = 1.9 \text{ m}^3/\text{sec} = 60 \text{ million m}^3/\text{month} \\ 1 & Q = 1{,}140{,}000 \Big/ (0.6 \times 9.81 \times 10 \times 24 \times 365) = 2.21 \text{ m}^3/\text{sec} = 70 \text{ million m}^3/\text{month} \\ 1 & Q = 1{,}635{,}000 \Big/ (0.6 \times 9.81 \times 10 \times 24 \times 365) = 3.17 \text{ m}^3/\text{sec} = 100 \text{ million m}^3/\text{month} \\ 0 & Q = 2{,}125{,}000 \Big/ (0.6 \times 9.81 \times 10 \times 24 \times 365) = 4.12 \text{ m}^3/\text{sec} = 130 \text{ million m}^3/\text{month} \end{cases}$$

The nonsymmetric Nash solution as explained in Chapter 2 is the unique optimal solution of the following problem:

$$\text{Maximize} \quad \prod_{i=1}^{4} (f_i - d_i)^{w_i} \tag{10.33}$$

$$\text{Subject to:} \quad d_i \le f_i \le f_i^* \quad \forall i \tag{10.34}$$

where w_i is the relative weight of objectives, f_i is the utility function, d_i is the disagreement point, and f_i^* is the ideal point of player (agency) i. In order to check the effect of the relative authority of these agencies (MATA), the symmetric Nash problem is also solved, and the optimal results are compared with the nonsymmetric results. Table 10.8 shows the results of the symmetric and nonsymmetric Nash problems. For this particular system, the annual water demand has been estimated

TABLE 10.8
Results of Symmetric and Nonsymmetric Nash Solutions for Example 10.9

	Nonsymmetric Solution (million m^3)		Symmetric Solution (million m^3)	
	Inflow = 1000	**Inflow = 1400**	**Inflow = 1000**	**Inflow = 1400**
Domestic	150	170	150	162
Agricultural	700	1030	700	1034
Industrial	80	130	80	130
Turbine Discharge (m3/sec)	2.21	2.21	2.21	2.32
Power Generation (MWh)	1140	1140	1140	1196

by an independent party to be at minimum about 110, 960, and 150 million m^3 and 1830 kWh for domestic, agricultural, industrial, and energy demands, respectively. In dry years (inflow = 1000 million m^3), domestic and agricultural demands are not supplied according to either solution; however, based on the symmetric solution, the energy demands have been supplied. In normal years, all demands have been supplied in both scenarios.

10.9 PROBLEMS

10.1 Monthly streamflow data for a gauging station are presented in the following table. Plot the flow–duration curve for this river.

Month	1999 (m^3/sec)	2000 (m^3/sec)	2001 (m^3/sec)
January	10	44	10
February	12	28	86
March	64	37	64
April	105	55	60
May	138	198	130
June	45	195	93
July	10	36	16
August	3	7	9
September	3	5	5
October	4	10	9
November	8	10	35
December	20	9	30

10.2 In Problem 10.1, a run-off river power plant is proposed at the site. The head available at the site is 50 m and plant efficiency is about 75%. Determine the expected firm energy at this plant.

10.3 In Problem 10.2, find the turbine design discharge if the turbines should run full 40% of the time.

10.4 Assume that a reservoir with storage of 800 million m^3 storage is built at the site explained in Problem 10.2. Firm release from the reservoir is considered to be 35 m^3/sec, which is also the turbine discharge capacity. The storage–water elevation relation at the reservoir site is given in the following table:

Elevation (m)	Storage (million m^3)
10	100
20	350
30	550
40	700
50	800

Use following relation for estimating the tailwater elevation:

$$\text{Tailwater elevation (m)} = \text{Release } (m^3/\text{sec}) \times 0.009$$

Assume that the reservoir storage at the beginning of the first month is 750 million m^3. Use the SSR method to estimated the expected monthly power generation in the reservoir.

10.5 In Problem 10.4, assume the reservoir should be operated based on the following rule curve. Find the monthly power generation of the reservoir and the firm energy.

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
E_t[a]	25	30	40	50	50	40	30	40	40	40	40	40

[a] E_t is the reservoir elevation at the beginning of month t(m).

10.6 For the system explained in Problem 10.4, consider that reservoir supplies irrigation demands of agricultural lands downstream of the reservoir as shown in the following table:

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
ID_t[a]	0	0	5	20	30	60	50	30	10	0	0	0

[a] ID_t is the irrigation demand in month t(10^6 m^3).

It also supplies part of the energy demands of the city, estimated to be 100,000 MWh per month. The extra power generated in this plant can be exported to other states. Formulate an optimization model for finding the optimal releases. Use the following cost functions for estimating the losses associated with shortages in supplying water demands and energy loads:

$$loss_t^1 = \begin{cases} 1.56 \times \left(D_t^W - R_t\right)^3 & \text{if} \quad D_t^W > R_t \\ 0 & \text{otherwise} \end{cases}$$

$$loss_t^2 = \begin{cases} 7.35 \times \left(D_t^L - P_t\right) & \text{if} \quad D_t^L > P_t \\ 0 & \text{otherwise} \end{cases}$$

where $loss_t^1$ and $loss_t^2$ are losses associated with water supply and power generation objectives, respectively. D_t^W and D_t^L are water and energy demands (in m^3/sec and MWh, respectively). R_t and P_t are water release and power generation, respectively, in month t.

10.7 In Problem 10.6, use the water and energy prices to formulate the optimization problem. Discuss the proper form of the objective function.

References

American Society of Civil Engineers, *Civil Engineering Guidelines for Planning and Designing Hydroelectric Developments*, Vol. 1, *Planning, Design of Dams, and Related Topics, and Environmental*, American Society of Civil Engineers, New York, 1989.

Creager, W. P. and Justin, J. D., *Hydropower Handbook*, John Wiley & Sons, New York, 1950.

Georgakakos, A. P., Yao, H., and Yu, Y., Control Model for Hydroelectric Energy-Value Optimization, *ASCE J. Water Resources Planning Manage.*, 123(1), 30–38, 1997.

Gulliver, J. S. and Arndt, R. E. A., *Hydropower Engineering Handbook*, McGraw-Hill, New York, 1991.

Karamouz, M., Energy and resource sustainability in developing countries, in *Proc. of the First Annual Conf. on Technological Advancement in Developing Countries*, Columbia University, New York, 1993.

Karamouz, M., Zahraie, B., Araghi-Nejad, Sh., A decision support system for monthly operating of hydropower reservoirs: A case study, *ASCE J. Computing Civil Eng.*, 2003 (Submitted).

Linsley, R. K., Franzini, J. B., Freyberg, D. L., and Tchobanoglous, G., *Water Resources Engineering*, McGraw-Hill, New York, 1992.

Mays, L. W., *Water Resources Handbook*, McGraw-Hill, New York, 1996.

Rangarajan, S., Surminski, H., and Simonovic, S. P., Reliability-based loss functions applied to Manitoba hydro energy generation system, *ASCE J. Water Resources Planning Manage.*, 125(1), 34–40, 1999.

Wong, K. P. and Wong, Y. W., Short-term hydrothermal scheduling. Part 1: Simulated annealing approach, *IEE Proc. Gener. Transm. Distrib.*, 141(5), 497–501, 1994.

Wood, A. J. and Wollenberg, B. F., *Power Generation, Operation, and Control*, John-Wiley & Sons, New York, 1996.

World Resources Institute, *World Resources: A Guide to the Global Environment: Toward Sustainable Development*, a joint publication by the World Resources Institute, The United Nations Environment Programme, The United Nations Development Programme, and The World Bank, Oxford University Press, London, 1992.

World Resources Institute, *World Resources 1998–99*, a joint publication by the World Resources Institute, The United Nations Environment Programme, The United Nations Development Programme, and The World Bank, Oxford University Press, London, 1998.

Yeh, W. W.-G. and Becker, L., Economic optimization of hydrothermal power systems operation, in *UERG Energy Science and Technology Report*, UER-114, University of California, Berkeley, CA, 1983.

Yeh, W. W.-G., Becker, L., Hua, S.-Q., Wen, D.-P., and Liu, J.-M., Optimization of real-time hydrothermal system operation, *ASCE J. Water Resources Planning Manage.*, 118(6), 636–653, 1992.

Yi, J., Labadie, W., and Stitt, S., Optimal unit commitment scheduling in hydropower systems, in *Proc. of ASCE International Conference of WaterPower '97*, 1997.

11 Water Demand Analysis and Management

11.1 INTRODUCTION

Water required for various purposes may come from groundwater or surface resources such as lakes, reservoirs, and rivers. It is called *untreated* or *raw* water and is usually transferred to a treatment plant. The degree of treatment depends on the raw water quality and the purpose for which that water will be used. Different water quality standards for domestic, industrial, and agricultural purposes are presented in Chapter 9. After treatment, the water usually enters a water distribution network. Details of the facilities required for water distribution are presented later in this chapter.

Providing an adequate water supply and sanitation to rapidly growing urban populations and developing agricultural and industrial activities is becoming increasingly difficult for different states throughout the world. The predominant approach to supplying these increasing water demands has been supply augmentation schemes. But, in many developed or developing countries, the costs of developing new sources or expanding existing sources are growing and becoming more physically and economically infeasible. The real costs of water per cubic meter in second- and third-generation projects in some cities have doubled between the first and the second project and then doubled again between the second and third (Bhatia and Falkenmark, 1993).

In recent years, more consideration has been given to saving water rather than developing new sources of water. In many countries, this approach has been accepted from both an economic and environmental point of view as being the best solution for meeting growing water demands. Therefore, water demand management is an appropriate strategy for improving efficiency and the sustainable use of water resources, taking into account economic, social, and environmental considerations (Wegelin-Schuringa, 2002).

The ability to manage existing water resources and then to plan for developing new water resources is directly tied to the ability to describe both present and future water use. The main objective of water demand management is to improve the efficiency and equity of water use and sanitation services. For this purpose, various instruments have been developed that can be generally classified into the following categories:

- Water conservation measures
- Economic measures
- Informational and educational measures
- Legal measures

The efficiency of each of these instruments depends greatly on local conditions. This chapter presents various aspects of managing water demands for domestic, industrial, and agricultural purposes.

11.2 WATER USE AND DEMAND: BASIC DEFINITIONS

Water demand is the scheduling of quantities that consumers use per unit of time for particular prices of water and is an analytical concept (Mays and Tung, 1992). Water use can be classified into two basic categories: consumptive and nonconsumptive. As shown in Table 11.1, domestic, agricultural, industrial, and mining water uses are consumptive. In contrast, hydropower, transportation, and recreation are the main nonconsumptive water uses, which are also *instream uses*. In this type of water

TABLE 11.1
Definitions of Water Use Terms

Term	Definition
Consumptive use	The part of water withdrawn that is evaporated, transpirated, incorporated into products or crops, consumed by humans or livestock, or otherwise removed from the immediate water environment
Conveyance loss	The quantity of water that is lost in transit from a pipe, canal, conduit, or ditch by leakage or evaporation
Delivery and release	The amount of water delivered to the point of use and the amount released after use
Instream use	Water that is used but not withdrawn from a ground- or surface-water source for such purposes as hydroelectric-power generation, navigation, water quality improvement, fish propagation, and recreation
Offstream use	Water withdrawn or derived from a ground- or surface-water source for public water supply, industry, irrigation, livestock, thermoelectric-power generation, and other uses
Public supply	Water withdrawn by public or private water suppliers and delivered to users
Return flow	The water that reaches a ground- or surface-water source after release from the point of use and thus becomes available for further use
Reclaimed wastewater	Wastewater treatment plant effluent that has been diverted for beneficial use before it reaches a natural waterway or aquifer
Self-supplied water	Water withdrawn from a surface or groundwater source by a user rather than being obtained from a public supply
Withdrawal	Water removed from the ground or delivered from a surface–water source for offstream use

Source: Mays, L. W., Ed., *Water Resources Handbook*, , McGraw-Hill, New York, 1996. With permission.

use, a use is made of a body of water without withdrawing water from it. *Onsite water use* refers to maintaining swamps and wetlands for wildlife habitats and ditching and ponding for soil management. The U.S. Geological Survey (USGS) definitions for ten water use accounts are given in Table 11.1. In the water-use circulars prepared by the USGS, the following water uses are identified within the overall water accounting system (Solley et al., 1993):

- Water withdrawal for *offstream* purposes
- Water deliveries at point of use or quantities released after use
- Consumptive use
- Conveyance loss
- Reclaimed wastewater
- Return flow
- Instream flow

Water use can also be classified as *offstream use* and *instream use*. Different water demands resulting from offstream uses can be classified as:

- *Domestic or municipal water uses* include residential (apartments and houses), commercial (stores and businesses), institutional (hospitals and schools), industrial, and other water uses (firefighting, swimming pools, park watering).
- *Industrial water uses* include water required for industrial processes such as cooling water for steam electric power generation units, refineries, chemical and still manufacturing, textiles, food processing, pulp and paper mills, and mining.
- *Agricultural water uses* include those for irrigating fields and for the drinking and care of animals.
- *Miscellaneous water uses* include fisheries, recreation, and mining.

These uses require withdrawal of water from the surface or groundwater resource system. Part of the water withdrawn may return to the system, perhaps at another location and time with different quality. The percent of return flows is an important factor in evaluating water use efficiency. It should also be considered in water resources management schemes. Table 11.2 shows a more detailed classification of water demands. As shown in this table, water use can be classified as municipal, agricultural, industrial, environmental, or infrastructure (public work).

Water demand changes from year to year and month to month. Many physical, economic, social, and political factors account for these variations. Significant climate changes have been observed in recent years in many parts of the world, producing more intense floods, greater precipitation, and even unusual droughts in some regions. Prediction of the demand variations in the future is the main objective in water demand forecasting, which is discussed in the next section of this chapter.

TABLE 11.2
Detailed Classification of Water Uses

Water Demand	Purpose Classification	Use-Type Classification	Rate of Consumption
Drinking	Municipal	Withdrawal	Low
Domestic (cooking, washing, etc.)	Municipal	Withdrawal	Low
Fish and wildlife	Agricultural and environmental	Withdrawal, instream, onsite	Moderate
Livestock	Agricultural, municipal	Withdrawal	Moderate
Drainage	Agricultural	Withdrawal, onsite	High
Irrigation	Agricultural, municipal	Withdrawal	High
Wetland habitat	Agricultural, environmental	Onsite	Moderate
Soil moisture conservation	Agricultural	Onsite	High
Utilization of estuaries	Agricultural	Instream, onsite	—
Recreation and water sports	Municipal, infrastructure	Instream	—
Aesthetic enjoyment	Municipal, infrastructure	Instream	—
Navigation	Infrastructure	Instream	—
Hydropower	Infrastructure	Instream	—
Mining	Industrial	Withdrawal	Moderate
Cooling	Industrial and municipal	Withdrawal	High
Boiling	Industrial and municipal	Withdrawal	High
Processing	Industrial and municipal	Withdrawal	Moderate
Steam power	Industrial and municipal	Withdrawal	High
Waste disposal	Industrial, municipal, agricultural	Instream	—

11.3 WATER DEMAND FORECASTING FOR REGIONAL AND NATIONAL PLANNING

Water demand should be forecasted in time and space. Many water resource projects have a relatively long useful life; therefore, in studies of regional water resources development, the time horizon for water demand forecasting should be extended to about 50 years. In medium-scale development plans, a lead time of 15 to 25 years may be applied. It should also be kept in mind that the forecast is made for general planning purposes for a relatively large region, rather than simply for construction of a single project. These projections should be made for at least three levels, namely normal, minimum, and maximum conditions.

The basic information required for long-term water demand forecasting includes projections of population and economic activities. The forecasts should reflect technological changes in production processes, product outputs, wastewater treatment

methods, social preferences, and public policies with respect to water use and development (Mays and Tung, 1992); otherwise, the forecasts would be of limited value to decision makers. The growing number of conflicts among water users is another important factor that should be incorporated in long-term water demand forecasting.

Because land, water, and other natural resources have region-specific characteristics, the response of industrial and agricultural activities to development policies would be expected to vary within and between regions. The following items indicate some of the main differences encountered among regions corresponding to the development of natural resources:

- Availability of natural resources
- Economic development
- Technological advancements in operation and utilization of natural resources
- Costs of water withdrawal
- Availability of labor skills
- Costs of transportation
- Natural resources markets
- Assimilative capacities of the environment

Water demand forecasts should reflect these regional characteristics. High water effluent standards and water prices may cause shifts in population. Within a long-term planning horizon, they may shift water demands from one region to another. In contrast, programs that support agricultural development (such as price-support programs) may encourage populations to move to regions offering more job opportunities. Water demand predictions based on simple extrapolation of current trends may lead to serious errors. The following data and information are necessary for forecasting the regional water demand over a long-term lead time:

- Population projections based on demographic studies and studies done by the agencies responsible for multiple-sector investment decisions
- Distribution of urban and rural population among subregions
- Gross national product (GNP)
- Projected outputs of agricultural, mining, electric power, and major manufacturing sectors for determining the regional distribution of activities based on the projected GNP
- Projected rates of per capita unit water use based on technological advancements and relative share of instream and offstream uses
- Expected use of brackish or ocean water

The water demand projections can be made without incorporating the water supply, at least at the first step. But, because these projections are likely to be revised if the costs of water are taken into account, a projection of demand must be accompanied by information about the supply curve of water for each planning region (see Chapter 4 for more information about the supply curve). Some of the water supply issues are presented in the following section.

11.4 WATER SUPPLY ISSUES

Four major characteristics of water supply are quantity, quality, time variation, and price. If the quantity and time variation of water resources conform with the water use in a region, then we do not need to store or regulate water. Otherwise, certain facilities should be implemented and the costs of the initial investment, operation, and maintenance should be incorporated in the supply curve. In the same way, if water quality does not satisfy the standards set for particular water uses, treatment plants should be implemented and their costs should be incorporated in water resources development studies.

In the previous chapters, large-scale and conventional methods of surface and groundwater resource development were discussed. Large-scale facilities such as dams, water transfer structures, and well fields are not always applicable to rural areas, small communities, and basins with limited surface and groundwater resources. In these cases, some unconventional methods should be considered for supplying water demands.

11.4.1 Rainwater Collection

Collecting rainfall is a primary or supplementary source at the household or small community level, especially in places with relatively high rainfall. Roofs of buildings are the most common collecting surfaces. Natural and artificial ground catchments are also used. When designing rainfall collection systems, the following issues should be considered (Falkland, 1991):

- *Quantity issues*: Rainwater collection systems often suffer from insufficient storage tank volumes or catchment areas. Leakage from tanks due to poor design, selection of materials, construction, or a combination of these factors is a major problem with rainwater collecting systems.
- *Quality issues*: The rainwater quality in many parts of the world is good, but in rainwater collection systems quality problems occur. Physical, chemical, and biological pollution of rainwater collection systems is found where improper materials have been used or where maintenance of roofs and other catchment surfaces, gutters, pipes, and tanks are lacking.

Figure 11.1 shows an example of a well-designed rainwater catchment system.

11.4.2 Desalination

Among the unconventional water supply methods, desalination is the most widely used. Seawater has a salinity of about 35,000 mg/l which is primarily composed of sodium chloride. Desalination of seawater in coastal areas or brackish water in inland areas is technically feasible, but it is expensive. The cost of desalination of seawater ranges from \$.80 to \$3/m^3 (Bouwer, 1994). In addition to a high rate of energy consumption, a major problem at inland desalination plants is the disposal of rejected brine. Evaporation ponds, injection in deep wells, and transfer to the ocean are common methods that depend on the volume of rejected brine, site location, and

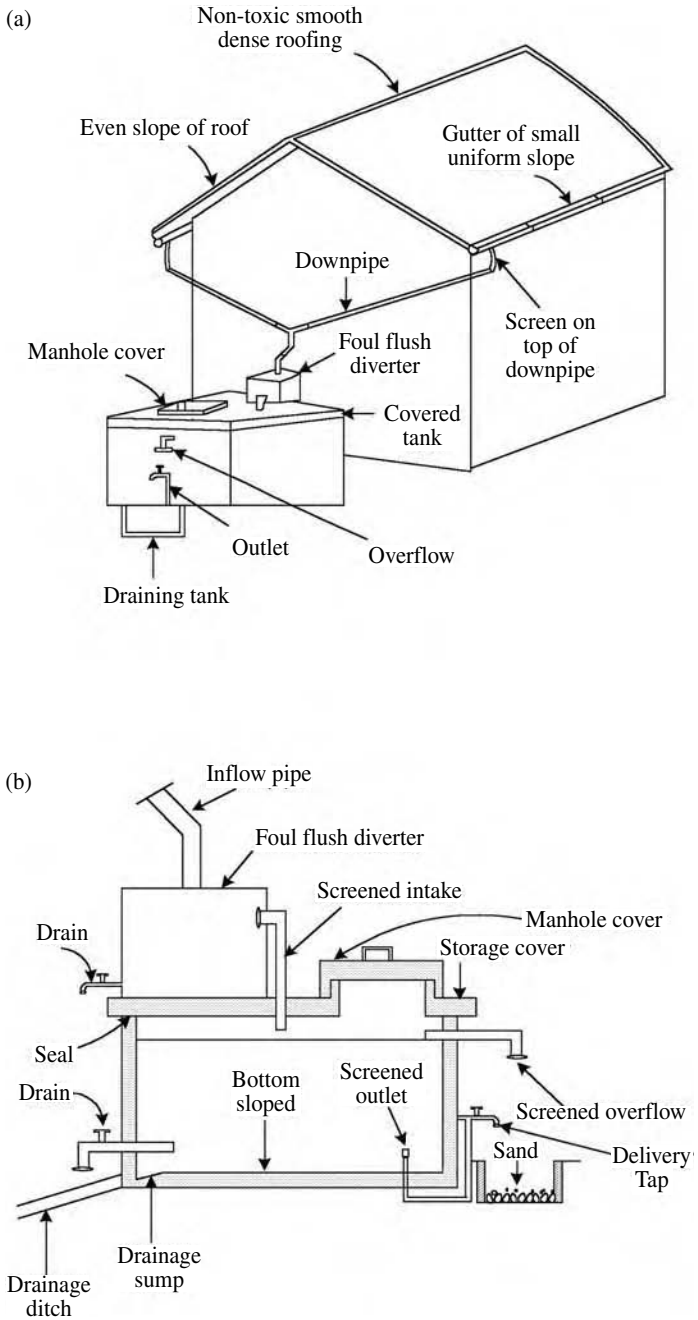


FIGURE 11.1 Examples of rainwater catchment systems. (From Falkland, A., Ed., *Hydrology and Water Resources of Small Islands: A Practice Guide*, United Nations Educational, Scientific, and Cultural Organization (UNESCO), New York, 1991.)

geographical and climatic conditions. The following methods are used for removing dissolved solids:

- *Distillation*: Water is heated to its boiling point to convert it into steam, then the steam is condensed to yield salt-free water.
- *Reverse osmosis* (*RO*): Water is forced through a semipermeable membrane under pressure, and the dissolved solids are held back.
- *Electrodialysis*: Ions are separated from the water by attraction through selective ion-permeable membranes using an electrical potential.

Among these methods, the first two are the most common processes for desalinization of seawater. Electrodialysis is usually preferred for brackish groundwater.

11.4.3 Artificial Recharge

The volumes of groundwater naturally replaced in each year are relatively small because of the slow rates of groundwater movement and the limited rate of infiltration. Artificial recharge can be used to reduce adverse groundwater conditions such as progressive lowering of water levels or saline water intrusion. Methods that have been used for artificial recharge include:

- *Utilization of holding basins*: Water is recharged by releasing it into basins formed by construction of dikes or levees or by excavation. Besides surface runoffs, local storms can also be diverted to these basins for artificial recharge.
- *Flooding*: This method can be used where topography is relatively flat so the flood can be spread over a large area. This method does not have land preparation costs.
- *Stream channel*: In this method, channels are modified to enhance the infiltration process. For this purpose, the time and area over which water is recharged should be increased. Upstream reservoirs will also help in regulating the streamflows based on the absorptive capacity of downstream channels.
- *Recharge well*: Recharge wells draw water from the surface to aquifers. Well recharging is practical where deep confined aquifers must be recharged or where economy of space, such as in urban areas, is an important consideration (Todd, 1980).
- *Furrow*: Water is distributed to a series of furrows. The furrows are usually shallow, flat bottomed, and closely spaced to provide maximum water contact area.

11.4.4 Groundwater Dams

Groundwater dams are subsurface dams constructed in aquifers. Their advantages over surface water storage dams are reduction in evaporation and improvement in water quality. This method has been used in Africa, India, and Cape Verde Islands (Hanson and Nilsson, 1986). Suitable geological condition must be present where

the water is stored. For this purpose, the rock where the dam is to be constructed should be relatively impermeable. Also, a suitable storage formation consisting of coarse sediments should be situated over the impermeable base. Further information on groundwater dams can be found in Nilsson (1988).

11.4.5 Weather Modification

One of the most widely used methods of weather modification is precipitation enhancement by *cloud seeding*, which requires an aircraft equipped with devices for distributing chemicals such as silver and iodide into clouds. Radar is usually used for detecting suitable clouds. Another method of weather modification is evaporation reduction, which is mainly used for small communities. Covering water tanks and deeper water storages (such as groundwater dams) can be used for evaporation reduction.

11.5 FLOW MEASUREMENT EQUIPMENT

Flow measurement is an important issue in water demand management. The water supply for various uses and demand points within a complex system, such as an urban area, is usually estimated based on flow measurement within the system. Illegal withdrawals and water losses can be detected by water measurement at different points in the water transfer and distribution system. Water measurement equipment includes those used in water networks (such as water meters) or those used in open channels or rivers for measuring flow flux (such as Parshal flumes or weirs).

11.5.1 Flow Measurement in Pipes

Three common methods are used to measure water flow in pipes. In the first method, differential pressure meters such as venturi tube, orifices, and nuzzles are used. In venturi tube meters, water flows through the tube and velocity is increased in the constricted portion, which temporarily lowers the static pressure in accordance with the energy equation. The pressure difference between the inlet and throat is measured and correlated to the rate of flow. Venturi meters are shaped to maintain streamline flow for minimum head loss.

The second method is based on measuring a chamber of known volume containing a disk that rotates as water passes through. The rotation resulting from the filling and emptying of the chamber is transmitted to a recording register. The advantages of this meter are simplicity of construction, high sensitivity and accuracy, small loss of head, and low maintenance costs. This oscillating disk meter is usually used for small customer services, such as individual households and apartments.

The third method used for measuring water flux in a pipeline utilizes a wheel with blades that rotates at a speed in proportion to the quantity of water that passes through the blades. A recording register is geared to the turbine wheel. The disadvantage of this method is poor accuracy at low flow rates when the water is not moving at sufficient velocity to rotate the blades. Consequently, current meters are used only in limited applications.

11.5.2 Flow Measurement in Open Channels and Rivers

Open channels are used for two purposes: (1) conveying water from water reservoirs for agricultural and other uses and (2) conveying wastewater. Wastewater contains suspended and floating solids that prohibit the use of enclosed meters. A Parshal flume is the device most commonly used to measure water flux in open channels. A typical flume consists of a converging and dropping open channel section. Flow moving freely through the unit can be calculated by measuring the upstream water level. A stilling well is normally provided to hold a float bubble tube or other depth-measuring device that is connected to a transmitter and flow recorder. The main advantage of an open channel flume is low head loss. Weirs can also be used for measuring flow of water in open channels. Water flowing over the sharp edge crest must discharge to the atmosphere and air must be allowed to pass freely under the jet. If these conditions are met, the rate of flow can be directly related to the height of water measured behind the weir. The most common type of weir used for measuring wastewater flows is the 90 V-notch weir. This kind of weir is usually installed on a temporary basis to make flow measurements associated with industrial wastewater surveys.

11.6 MUNICIPAL WATER DEMAND

As mentioned previously, municipal water uses include residential (apartments and houses), commercial (stores and businesses), institutional (hospitals and schools), industrial, and other water uses (firefighting, swimming pools, park watering). Municipal water use is usually quantified as the volume of water used by each person in each day, referred to as *per-capita water use.* It is usually measured in liters per day or gallons per day per person or per metered service.

Globally, variation in the withdrawal of water for municipal purposes varies enormously. Average municipal water use in the United States per metered service is about 2270 l/day. Water consumption of residential areas in eastern and southern states is about 790 l/day; in central states, about 1060 l/day; and in western states, about 1740 l/day (Hammer and Hammer, 1996). Lawn sprinkling can have a significant influence on water demand in areas with large residential lots. In these areas, 50 to 75% of the total daily volume may be attributed to irrigation. Figure 11.2 shows per-capita water use in some of the big cities in the world.

Municipal water demand depends greatly on health standards in urban and rural areas. The quality of the drinking water has a direct effect on health, and the amount and quality of water available for bathing and cleaning have a significant indirect impact on health. Per-capita water use in most residential areas has increased over the past decades. Most new houses have more water fixtures, spacious lawns, modern appliances, and other conveniences that consume larger volumes of water.

11.7 MUNICIPAL WATER DEMAND ESTIMATION AND FORECASTING

Costs of water supply services and technological developments designed to lower these costs have a major influence on the level of water demand in developing

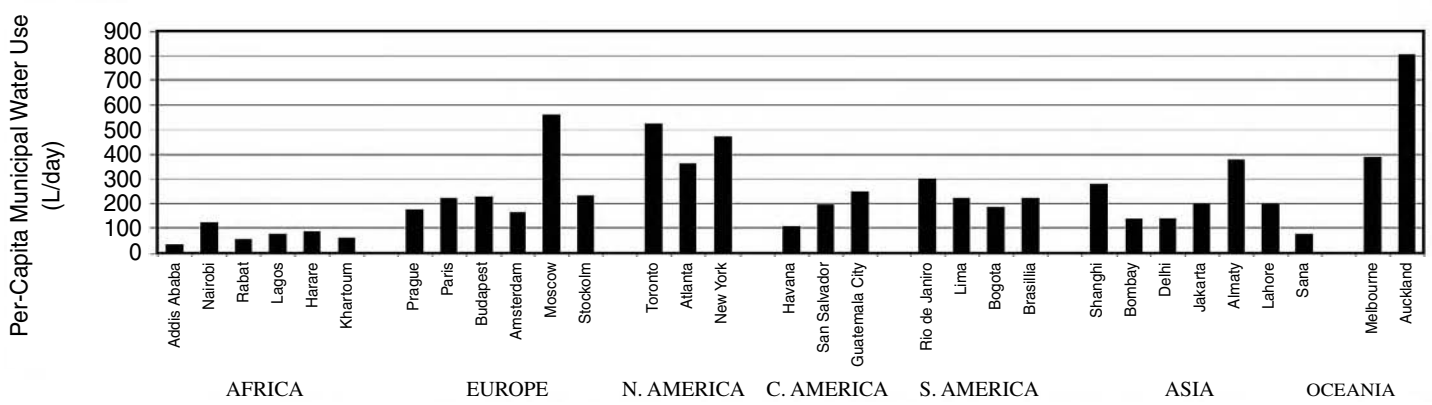


FIGURE 11.2 Per-capita water use in large cities of the world. (Data tables from World Resources Institute, 1998.)

countries. In rural areas, the distance between standpipes or the number of persons served by a single tap or well are decisive factors controlling the level of demand. Precise estimation of municipal water use can be obtained by breaking down the total delivery of water to urban areas into a number of classes of water use and determining separate average rates of water use for each class. This method is referred to as *disaggregate estimation of water use*. Previous studies have shown that water use within some homogenous sectors is less variable compared to the total water use; therefore, greater accuracy in estimation of water use can be obtained. The components of urban water usage can be classified as follows:

- Domestic
  - Washing and cooking
  - Toilet
  - Bath and/or shower
  - Laundry
  - House cleaning
  - Yard irrigation
  - Swimming pool
  - Car washing
  - Other personal uses (hobbies, etc.)
- Public services
  - Public swimming pools
  - Governmental agencies and private firms
  - Educational services (such as schools, universities, and dormitories)
  - Firefighting
  - Irrigation of parks, golf courses, etc.
  - Health services (such as hospitals)
  - Public baths, public toilets, etc.
  - Cultural public services (such as libraries and museums)
  - Street cleaning and sewer washing
  - Entertainment and sport complexes (such as cinemas and clubs)
  - Food and beverage services
  - Accommodation services
  - Barber shops and beauty parlors
- Small industries (such as laundries or workshops)
- Construction and public works
- Water losses
  - Leakage from pipes, valves, meters, etc.
  - Evaporation in open reservoirs
  - Overflow of reservoirs
  - Disrepair of elements of a water distribution network, such as cracked reservoirs, flow back through one-way valves and pumps, etc.
  - Loss in production process (cooling, pumping, etc.)
- Transportation
  - Taxies, buses, and other conveyances (stations, garages, etc.)
  - Ports and airports
  - Railways (stations and workshops)

In order to estimate the total municipal water use in a city, the study area should first be divided into homogenous areas. These homogenous areas are usually selected based on pressure districts or land-use units, and the water use rates can be assumed to be constant for different users within each area. Temporal (e.g., annual, seasonal, monthly) variation may also be considered when subdividing the water uses.

The most commonly used method to forecast water use is regression analysis. A simple regression model can be formulated as follows:

$$Q_{t,i,j} = aX_{t,i,j} + b + \varepsilon_t \tag{11.1}$$

where:

$Q_{t,i,j}$ is the average rate of water use in time period t for disaggregated use i in homogenous subarea j.
$X_{t,i,j}$ is the independent variable in time period t for disaggregated use i in homogenous subarea j.
ε_t is the error term in time period t.
a and b are regression coefficients.

$Q_{t,i,j}$ is the dependent variable of the regression. The independent variable, $X_{t,i,j}$, should be selected based on the available data of different factors affecting the water use and the relative importance of them in increasing or decreasing water uses. For example, the most important factor in estimating water use in an urban area is the population in each subarea. In Eq. (11.1), if the independent variable is set to be equal to the number of users in each subarea, then the regression slope coefficient, a, would be the per-capita water use. The water use can also be estimated based on the number of connections to the water distribution system as follows:

$$Q_{t,i,j} = e \cdot NC_{i,j} \cdot W_{t,i,j} \tag{11.2}$$

where:

$NC_{i,j}$ is the number of connection for disaggregated use i in homogenous subarea j.
$W_{t,i,j}$ is the water use per connection for disaggregated use i in homogenous subarea j in time period t.
e is the efficiency of the water distribution system, which is a function of leakage and other water losses in the system.

Multiple regression methods can also be used to incorporate other variables correlated with water use in municipal areas to estimate and forecast water use in the future. Population, price, income, temperature, and precipitation are some of the variables that have been used. Based on the independent variables, water use models can be classified as *requirement models* and *demand models*. In requirement models, only the physical and ecological variables related to water use are incorporated, but

demand models are based on economic reasoning and include only variables that are correlated significantly with water use and are expected to be causally related to water use (Mays and Tung, 1992). Linear and logarithmic water use models have been suggested by different investigators:

$$Q = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n + \varepsilon \tag{11.3}$$

$$Q = a_1 \ln x_1 + a_2 \ln x_2 + \cdots + a_n \ln x_n + \varepsilon \tag{11.4}$$

Besides regression models, time series analysis has also been used to forecast future variations of water demands. For this purpose, time series of municipal water use and related variables are used to model historical patterns of variations in water demand. Long memory components, seasonal and nonseasonal variations, jumps, and outlier data should be carefully identified and used to model water demand time series. Details of time series analysis and statistical modeling are provided in Chapter 5.

11.8 ELEMENTS OF URBAN WATER DISTRIBUTION NETWORKS

Water distribution networks usually consist of the following elements:

- The *reservoir* provides water to the water distribution system. When modeling water distribution systems, a reservoir is usually considered as a boundary node with constant surface water elevation. In other words, it is assumed that inflow and discharge from the reservoir do not affect its surface water elevation. Important information about reservoirs includes the surface water elevation and water quality.
- The *tank* stores excess water within the water distribution system and releases it when the water use is high. Tanks are similar to reservoirs but their storage capacity is smaller, and the hydraulic grade line in tanks varies with respect to inflow and outflow variations.
- *Pipes* convey water from one point to the other. Pipes used in distributing water under pressure include ductile iron, plastic, concrete, and steel. Copper and plastic pipes with small diameters are usually used for house connections. Pipes in water distribution networks should have enough tensile and bending strength to withstand external loads due to trench backfill and earth movement caused by freezing, thawing, or unstable soil conditions. They should also have enough bursting strength to withstand internal water pressures and be able to resist impact loads encountered in transportation, handing, and installation. The inner side of the pipes should be smooth and noncorrosive to create minimum resistance to water flow (Hammer and Hammer, 1996). Two important characteristics of pipes are length and diameter. When modeling water distribution networks, the length assigned to a pipe should represent the total distance between two

nodes and not necessarily the straight-line distance. Pipes are usually produced in lengths of around 6 m and are referred to by their nominal diameter. The internal diameter of the pipes, which should be used in modeling, is usually different from the nominal diameter. Internal diameters may change over time due to tuberculation, corrosion, etc.

- A *junction* refers to any of the following nodes in a water distribution network:
  - Location where two or more pipes meet
  - End of a single pipe (dead-end pipe)
  - Location where water is withdrawn from the network
  - Location where water is injected into the network
- *Pumps* are used for a number of purposes in water distribution networks; however, they are usually used to boost the head at specific locations to overcome piping head losses and physical elevation differences. Some of the main types of the pumps can be classified as:
  - *Low-lift pumps* elevate water from a source to a treatment plant.
  - *High-service pumps* discharge water under pressure to a distribution system.
  - *Booster pumps* increase pressure in a water distribution system.
  - *Circulation pumps* move water in a treatment plant.
  - *Well pumps* lift water from wells for water supply.
- In water distribution networks, the most commonly used type of pump is the *centrifugal pump*. A centrifugal pump has a motor that spins a piece within the pump called an *impeller*. The mechanical energy of the rotating impeller is imparted to the water, resulting in an increase in head.
- A *valve* is an element for controlling either the magnitude or direction of water flow through a pipeline. Some of the main types of the valves can be classified as:
  - *Shutoff valves*, the most common types of the valves in water distribution systems, are used to stop the flow of water through a pipeline. The primary purpose of this type of valve is to turn off a portion of the system, for example, for the time it takes to replace a broken pipe. Different types of shutoff valves include gate, butterfly, globe, and plug valves. Among all these types of shutoff valves, gate valves are the most widely used valve.
  - *Check valves* are used to ensure that water flows only in one direction in the pipeline. Water flowing backward through the valve causes it to close. They are usually installed in the discharge pipes of centrifugal pumps to prevent backflow when the pump is off.
  - *Altitude valves* are used to automatically control the flow into and out of an elevated storage tank or standpipe to maintain desired water-level elevation. When the tank level rises to a specific upper limit, the valve closes to prevent any further flow from entering.
  - *Air release valves* release trapped air during the operation of water distribution systems. Air can enter a pipe network through leaking joints, from pipes drawing air into the suction pipes, etc. These air

pockets increase resistance to the flow of water. Air release valves can be used at these locations to discharge the trapped air.
- *Control valves* can be classified as flow control valve, throttle control valve, and pressure-based control valves. Pressure-based control valves are designed to maintain the hydraulic grade or the pressure at a specific level.

11.9 MUNICIPAL WATER DEMAND MANAGEMENT MEASURES

The primary measures of domestic water demand management can be classified as:

- Water loss reduction:
  - Leakage detection and reduction
  - In-house retrofitting
  - Reduction of illegal connections
- Education and training:
  - Public awareness
  - In-school education
  - Training and education of the staff in water related agencies
- Economic incentives and water pricing:
  - Water metering
  - Tariff structure
- Institutional measurements and effective legislation:
  - Regulations for water demand management
  - Regulations on resale of water
- Water reuse

A brief explanation is presented here about each of the above water management measures.

11.9.1 Water Loss Reduction

Throughout the world, a large amount of water is lost through urban water supply systems due to leakage in distribution networks and home appliances or through illegal connections. These water losses are referred to as *unaccounted-for water*, a measure that is often used to quantify the efficiency of a water supply system. Unaccounted-for water can be defined as the difference between the amount of water supplied from the water works, as measured through its meters, and the total amount of accounted-for water. Accounted-for water includes water consumption as recorded by customers' meters, water stored in service reservoirs, and authorized free use such as for flushing and sterilization of mains and routine cleaning of service reservoirs. In many cities unaccounted-for water is estimated to be as high as 50% of water supplied. A certain level of unaccounted-for water cannot be avoided, but in some cities such as Singapore the level has been reduced to about 6% (United Nations, 1998). Unaccounted-for water can be classified as:

- *Apparent water loss*, which includes errors due to meter inaccuracy and improper accounting of water used in the commissioning and filling of new mains, connections, and service reservoirs and for cleaning and flushing of the water distribution system during maintenance
- *Real water loss*, which includes water loss due to leaks and illegal drawoffs from the transmission and distribution system

The following items can be considered as control measures for reducing the amount of unaccounted-for water:

- *Leakage detection*. A specific program is required to detect the leakage. It can include visual inspection for leaks along transmission and distribution pipelines and also leak detection tests at night for distribution mains. Nighttime is selected because the pressure is usually higher in the system, making it is easier to detect the leaks and also minimizing inconvenience to customers. To locate the leaks, mechanical, electronic, and computerized acoustic instruments such as stethoscopes, geophones, electronic leak detectors, and leak noise correlators can be used. The time frequency of testing is important. It depends on the percentage of unaccounted-for water and share of leakage and the size of the water distribution systems.
- *Leakage control*. Better quality pipes and fittings can be used for leakage control. More durable and corrosion-resistant pipe materials such as ductile iron pipes internally lined with cement mortar, copper, and stainless steel can be used for this purpose. When designing the water distribution network, minimizing the number of joints is important to reduce minor leaks. Teflon packing for repairing valve glands and leaks due to wear and tear and dezincification-resistant brass fittings can also reduce the minor leakage in the water distribution network.
- *Full and accurate metering policy.* Meters with less accuracy such as Venturi/Dall tubes can be replaced with more accurate meters such as electromagnetic meters. A specific program for maintaining and replacing meters regularly should also be considered. This can be done through bulk changing programs.
- *Proper accounting of water used*. In most water distribution networks in urban areas, significant quantities of water are used in the commissioning and filling of new mains, for connection and service reservoirs, for cleaning and flushing the distribution system during maintenance, and for firefighting. Water used for such purposes should be accurately reported to ensure proper accounting.
- *Strict legislation on illegal drawoff*. The illegal drawoff should be detected and anyone who is responsible for it should be prosecuted.
- *Education and training*. Public education and training are important issues in achieving the water demand management goals. Various methods have been tried in developed and developing countries such as:
  - Training programs in the mass media
  - Billboards along streets and on public transport vehicles

- Training brochures enclosed with water bills

  Educating school children has been given top priority in many countries. These programs present basic information about the hydrologic cycle, costs of water supply, maintenance of water distribution networks, water pollution, and costs of wastewater treatment. Education and training can also be of benefit to other water management strategies, such as changing water tariffs and charging effluents. The staffs of water-related agencies should be well trained. Training staff to be technically competent and customer service oriented should be considered an important objective when designing staff training programs.

11.9.2 Economic Incentives and Water Pricing

Pricing is a powerful tool for controlling urban water use. In many countries, water and water-related services have been provided almost free or at a very low charge, especially to domestic customers. Water charges in many urban areas do not even cover the cost of operation and maintenance of a water supply system; therefore, those services have been highly subsidized. When defining tariffs for urban areas, groups with high, medium, and low levels of income and the areas in which they live should be identified. Various structures of tariffs can then be applied to create relative equity in the allocation of the costs of water services and to control the rate of domestic water use. According to the United Nations (1998), increased tariff rates can be of benefit if:

- They reflect the real costs of the services provided.
- They are linked to consumption levels.
- Differential charge increases are large enough to encourage water savings.
- Any changes in the tariff rates are accompanied by well-organized public information campaigns.

11.9.3 Institutional Measurements and Effective Legislation

Fragmentation and overlapping of responsibilities among agencies involved in water resources management are obstacles to efficient water use in urban areas. In most countries, different organizations are responsible for the following tasks:

- Water resources development planning
- Water supply to urban water distribution network
- Wastewater collection
- Water pollution control

The levels of responsibility for these agencies may also differ at national or more limited provincial or city levels. These tasks may be handled by a number of organizations, and the decisions made in each of them can significantly affect other responsible agencies. In addition, the urban sector is closely interrelated with other economic sectors, which is why any proposed program for urban water conservation will not be successful if it attempts to isolate the water supply and sanitation sector

from other urban sectors. Thus, integrated management policies are required to achieve economical use of water in urban areas (United Nations, 1998); however, such policies can only be put into practice with the commitment and close cooperation of all the organizations and agencies involved. For this purpose, conflict resolution issues should be incorporated in the development of planning and management policies.

The efficiency of implementing water resources development projects and utility management can be improved by allowing the private sector to manage and handle the associated risk, but the governmental agencies should maintain overall control through legislation and regulation. The injection of private-sector efficiency and productivity also results in reduced public-sector expenditure.

11.9.4 Water Reuse

Details of water reuse for different purposes are presented in Section 11.16.

11.10 AGRICULTURAL WATER DEMAND

Water use in agriculture is continuing to increase as world food demand rises. Agriculture already accounts for about 70% of water consumption worldwide. The United Nations projects up to a 100% increase in irrigation water by the year 2025 (World Resources Institute, 1998) and forecasts that nations will develop new sources of water. In Asia, Central America, and Africa, the share of agricultural water use is about 85 to 90% of all water uses; therefore, water demand management issues in the agricultural sector can play a significant role in reducing water demands.

Crop water demand is usually measured in terms of evapotranspiration. The rate of evapotranspiration depends on meteorological factors such as relative humidity, temperature, wind, radiation, number of daylight hours, precipitation, etc. and field factors such as soil moisture and surface properties. Crop water requirements are usually estimated based on the rate of potential evapotranspiration, ET_o. Potential evapotranspiration is defined as the maximum evapotranspiration rate that the atmosphere is capable of extracting from a well-watered field. Methods for estimating potential evapotranspiration have been developed by Blaney and Criddle (1945), Thornthwaite and Mather (1955), Penman (1948), and Thomas (1981). Details of the Thornthwaite method are presented in Chapter 12. The crop water requirement can then be estimated as follows:

$$ET_{crop} = K_c \cdot ET_o \tag{11.5}$$

where ET_{crop} is the water requirement for a specific crop, and K_c is the empirical crop coefficient, which depends on the following factors:

- Climate
  - Temperature
  - Wind
  - Humidity

- Crop
  - Stage of growth
  - Crop reflectivity
  - Crop height and roughness
  - Degree of ground cover
  - Canopy resistance to transpiration

For most crops, the K_c value varies from 0.6 to 0.9 for the total growing season. In order to estimate the agricultural water demands, the following items should be incorporated:

- *Crop water requirement* is the amount of water that crops require to grow and is considered to be equal to evapotranspiration.
- *Irrigation requirement* is the crop water requirement minus effective rainfall available for plant growth. In estimating the irrigation requirements, water consumed for leaching and other miscellaneous requirements such as for germination, frost protection, and plant cooling should be included, and the soil moisture loss should be subtracted.
- *Farm delivery requirement* is estimated as the irrigation requirement plus farm water losses due to evaporation, deep percolation, surface waste, and nonproductive consumption.
- *Gross water requirement* is estimated based on farm delivery requirements plus seepage losses in the canal system from the headworks to the farm unit plus the waste of water due to poor operation.

Agricultural water demands are changed from year to year and month to month. The parameters affecting agricultural water demands can be summarized as follows:

- *Crop mix*. Each crop has its own water demands; therefore, the types of crops and the areas covered by each crop, which usually vary from year to year, can dramatically change the agricultural water demands of a specific farm.
- *Irrigated and dry-land farming*. Substitution of irrigated farming for dry-land farming frequently occurs in areas with a sufficient rate of effective rainfall.
- *Period and sequence of cropping*. The time of planting may also change the water demands of crops. These variations are highly dependent on climatic characteristics.
- *Physical characteristics of the water transfer and irrigation systems*. The physical characteristics of irrigation systems are important in estimating water losses through a transfer and irrigation system. The *water supply efficiency* can be defined based on the project, field, and farm efficiency. *Project efficiency* can be estimated based on the proportion of flow at the source and water applied to crops. *Field efficiency* deals with the proportion of water flowing into the field to the amount of water that is applied

to crops. *Farm efficiency* can be estimated based on the proportion of water at the turnout to the amount of water that is applied to the crops. In addition to water supply efficiency, *irrigation efficiency* can be defined in terms of financial, physiological, and sociological aspects. *Financial efficiency* can be defined as the financial return for a particular amount of investment in water supply and can vary significantly from year to year and from place to place. It is not always possible to quantify the long-term sociological benefits of irrigation. In drought events, where the indices of drought are not predictable, even a modest contribution to food security arising from assured irrigation can have very profound consequences on the quality of life of the people.

- *Market prices*. A change in the market prices of agricultural products is another important issue in changing the agricultural water demands. Farmers quickly respond to market changes due to economic incentives.
- *Climate variation*. Hot weather, a windy climate, and low humidity are factors that increase crop water requirements. Also, the net effective rainfall supplies part of the crop water demands, so crop irrigation demands will be less in years with greater rainfall.
- *Policies related to pricing, importing, and exporting of agricultural products*. National policies about importing and exporting agricultural products can drastically change product prices. Agricultural price-support programs have significant effects on increasing agricultural water requirements.

11.11 IRRIGATION METHODS

Increasing irrigation efficiency is the most important issue in agricultural water demand management; therefore, it is important to recognize various methods of irrigation and their advantages and disadvantages. The irrigation methods can be classified as:

- *Gravity or surface irrigation*. Water is delivered to farms via canals and then distributed to individual fields by gravity whenever possible. Pumping, if required, increases operating costs proportionally. Surface irrigation includes furrow, flood, and strip methods. Furrow irrigation is used on crops grown in rows, with the water confined to running laterally between the raised rows of crops. Flood irrigation is used on close-growing field crops (such as rice) where the field or part of the field is surrounded by shallow levees so that the entire field can be flooded. For the surface irrigation method, the soil is the medium through which water is conveyed, distributed, and infiltrated. This method is mechanically simple, has low energy requirements, and can be easily adopted to small holdings. Its serious deficiencies are low application frequency, high conveyance losses, and wastage of water, which lead to adverse consequences such as water logging and salinization. Figure 11.3 shows examples of flood and furrow irrigation systems.

(a)

(b)

FIGURE 11.3 Examples of (a) flood and (b) furrow irrigation systems.

- *Sprinkler irrigation systems*: In these systems, sprayed water falls on plants like rain. These systems are designed either to be portable and move across a field with the water supply hose attached by cable (*traveling sprinklers*) or to rotate continuously around a pivot point (*center pivot systems*). Sprinkler systems have the advantage of providing a more controlled application rate than surface irrigation systems. The conveyance losses are minimal, and it allows soil aeration. The main disadvantages of this method are high capital requirements and operation and maintenance costs, as well as high sensitivity to wind. Figure 11.4 shows examples of sprinkler irrigation systems.
- *Drip irrigation*: In this method, water is applied to soils directly under the surface; to the root zone. Water is delivered drop by drop through perforation or emitters in the pipes. The trickling rate is maintained at a level less than the rate of infiltrability of soil. This method can be used for fields for which the upper subsoil is very permeable with uniform topography and moderate slopes. The advantages of this method are low labor costs and high crop yields. The capital costs of drip irrigation systems are high. The method requires rigorous adherence to scheduling and maintenance. A microprocessor-based drip irrigation system suitable for use in developing countries is now commercially available. This device uses low-cost ceramic sensors and operates on a solar-cell battery. The device continuously monitors the soil moisture and controls the drip rate so as to maintain the moisture within the desired limits. Figure 11.5 shows examples of drip irrigation.
- *Microsprayer irrigation systems*. In these systems, the water is applied only to a fraction of the ground surface. The advantages of these systems are similar to those for drip irrigation systems, including high-frequency irrigation and allowing injection of fertilizers, if desired. Microsprayer irrigation systems also have the following advantages over drip irrigation systems:
  - As microsprayers have larger nozzle orifices, clogging of emitters is not a serious problem, and it is not necessary to filter the irrigation water.
  - They are operated at a pressure on the order of 2 atmospheres, a level that is much lower than for the drip systems.
- They can be scaled down for use in small farms.

The minor disadvantages of this method include:

- Because the area wetted is larger than that in drip systems, some evaporation losses occur.
- Because the leaves get wet, brackish water cannot be used in irrigation.
- *Low-head irrigation*: In a closed conduit system, the conveyance losses are avoided but the system requires energy to pressurize water for distribution. Low-head systems have the advantages of a piped system but do not require any pumps or nozzles. In these systems, water is simply

(a)

(b)

(c)

FIGURE 11.4 Different types of sprinkler irrigation: (a) center pivot system; (b) sprinkler irrigation; (c) wheel line field irrigation.

FIGURE 11.5 Examples of drip irrigation.

allowed to bubble out from open, vertical standpipes that rise from buried lateral irrigation tubes. These systems are often used for flow rates typically on the order of 30 to 200 l/sec and are particularly suitable for widely spaced crops such as fruits trees and grapevines. Low-volume, high-frequency, partial-area irrigation is possible with this system. The initial cost of the system is comparable to or even lower than other systems, it incurs no energy costs, and the useful life of the system is longer.

11.12 DRAINAGE AND SALINITY CONTROL

Drainage refers to artificial removal of excess water from agricultural lands. Drainage may be of two kinds: surface drainage and groundwater drainage. Surface drainage refers to shaping the land so as to facilitate overland flow, which drains excess water accumulation on the surface. Lowering the groundwater table or preventing its rise (groundwater drainage) drains excess water within the soil or subsoil.

While soil saturation for short periods is not harmful by itself, prolonged saturation (water logging) can adversely affect plant growth. Excess water in the soil blocks the soil pores, impedes the movement of oxygen from the atmosphere, and limits the respiration activity of the plants. A decrease in the solubility of oxygen in water and an increase in the respiration rate of both plants and microorganisms result in an increase in temperature. Thus, water logging has more severe consequences in warm climates than in cold climates. The land becomes infertile due to the steady rise of the water table, water logging, and salinization. Thus, all irrigated lands must have a drainage system, and irrigation without drainage can have disastrous consequences.

Excess water from the soil or subsoil is drained through ditches, perforated pipes, or machine-formed mole channels by gravity flow or pumping. The drained water is transferred into a stream, lake, or an evaporation pond. The rate of flow from the soil to the drains depends upon the following factors:

- Permeability of the soil
- Depth of the water table
- Depth of the drain
- Horizontal spacing between the drains
- Configuration of the drains (open ditches or tubes) and their diameter and slope
- Nature of the drain-enveloping material (such as gravel) used to increase the seepage surface
- Rate of recharge of groundwater

As the field conditions tend to be highly complex and variable, empirical equations must be used to estimate the desirable depths and spacing of drain pipes under particular soil and groundwater conditions.

11.13 AGRICULTURAL WATER DEMAND MANAGEMENT

The agricultural water demand management measures can be classified as:

- Increasing irrigation efficiency
- Development of proper crop mix
- Soil development and management
- Wastewater reuse in irrigation

11.13.1 Increasing Irrigation Efficiency

The types of irrigation systems were explained in the previous section. When selecting the proper irrigation system for a specific district, the following considerations are important:

- Physical site condition
- Social and institutional considerations
- Economic considerations

General site conditions as listed in Table 11.3 should be considered when selecting the proper irrigation system. In addition to the general classification of irrigation systems, the following items also affect the irrigation efficiency:

TABLE 11.3
Physical Site Conditions to Consider in Irrigation System Selection

Crops	Land	Climate	Water Supply	Energy
Crops grown	Field shape	Precipitation	Source	Availability
Crop rotation	Obstructions	Temperature	Quantity	Reliability
Crop height	Topography	Frost condition	Quality	Cost
Cultural practices	Soil	Humidity	Salinity	
Disease potential	Texture	Wind	Sodicity	
Pests	Uniformity		Sediments	
Water requirements	Depth		Organics	
Climate modification	Intake rate		Reliability	
	Water-holding capacity		Delivery	
	Erodibility		Schedule	
	Salinity		Frequency	
	Drainability		Rate	
	Bearing strength		Duration	
	Flood hazard		Water cost	
	Water table			

Source: Mays, L. W., Ed., *Water Resources Handbook*, McGraw-Hill, New York, 1996. With permission.

- Appropriate on-farm facilities (measuring devices, regulatory structures, ditches) ensure good managerial control over the delivery and distribution system. For this purpose, training programs for the proper use of the facilities are of paramount importance.
- Loss of water by direct evaporation from the soil surface is a significant source of water loss and inefficient irrigation. Improved irrigation technologies should be used to reduce evaporation losses.
- A preference for using groundwater as a source of agricultural water supply reduces losses at the source, particularly in regions with a hot climate and high evaporation rate. If surface water is the main source of water supply, reduction of evaporation and seepage losses from reservoirs should be considered.
- Automated control and operation of irrigation systems greatly contribute to increasing the efficiency of water use by preventing or reducing losses due to managerial failures and over-irrigation.
- When farmers are confident water will be available when needed, they are less likely to irrigate to excess. Farmers frequently over-irrigate to ensure that enough water is applied to last until the next time water is available. Delivery systems in which water is always available and notification is not required are referred to as *on-demand delivery systems*.
- Legal issues may also affect the efficiency of water use in irrigation. Water rights are the most important issue related to the legal system. Transfer restrictions on water use rights may prevent the utilization of water on more productive lands or for more productive purposes. Quality criteria are not usually specified in water use rights; however, they have a significant impact on the efficiency of water use. Water rights are sometimes issued on a first-come, first served basis, which may tend to prevent or restrict the efficient use of water in terms of larger areas and longer periods.
- Pricing is another important issue in agricultural water demand management. Some districts charge a fixed rate based on the land area served, regardless of the amount of water delivered. Others charge for the volume of water delivered, and the price for unit volume is usually defined based on the price of agricultural products in each field. This encourages conservation but makes the financial management more difficult.

11.13.2 Development of the Proper Crop Mix

In many irrigation systems, the water is supplied from local rivers and streams with no or insufficient regulation. In such cases, on-time use of available surface runoffs may be an important objective in increasing productivity of agricultural lands. For this purpose, an optimal crop mix can be developed. In order to develop an optimal mixture of crops for specific agricultural land, an optimization model can be developed. In the following formulation for this model, on-time use of available water is maximized considering monthly variations of demand and streamflows. The objective function of this model is as follows:

$$\text{Minimize} \quad z = \begin{cases} \sum_{j=1}^{N} \sum_{i=1}^{12} (I_{ij} - D_{ij}) & \text{if} \quad I_{ij} \geq D_{ij} \\ 0 & \text{otherwise} \end{cases} \tag{11.5}$$

where:

N is the number of crops.
i is the number of months ($i = 1, \ldots, 12$).
I_{ij} the volume of water allocated to crop j in month i.
D_{ij} is the volume of water demand of crop j in month i.

This objective function will minimize losses for months in which the surface runoffs are not allocated. To limit possible shortages when supplying irrigation demands, constraints can be considered to limit the maximum shortage rates in different months. Agencies responsible for irrigation water supply usually define tariffs based on the price of agricultural products; therefore, water allocated to agricultural lands may be priced differently. The difference in price may also be incorporated in finding the optimal crop mix for each specific region.

11.13.3 Soil Development and Management

Suitability of lands for irrigation is also an important issue in increasing irrigation efficiency and productivity of agricultural lands. For this purpose, relocation of agricultural activities in more cultivated lands, if possible, can be considered as an option at the local and regional agricultural planning level.

11.14 INDUSTRIAL WATER DEMAND MANAGEMENT

In developing countries, industrial water demands generally have been rising, but in other countries such as Japan the increasing rate of water demand has slowed down or even reversed. Technological advancements in such industries as pulp and paper production can significantly reduce the water demands. Development of recycling techniques also has played a vital role in reducing industrial water demand in recent years. Recycling water in cooling systems has been initiated in many countries, and water reuse has become an important issue in industrial water demand management. Water quality standards for many industries can be achieved by conventional treatment methods. Domestic and industrial wastewater can be treated and reused in industrial processes. Water recycling and reuse are also appropriate options for reducing the costs of wastewater treatment before discharge to water bodies.

Economic incentives can also be effectively used for controlling water-intensive and water-polluting industries such as steel and petrochemicals and textiles. Water tariffs, effluent charges, and tax concessions for the use of water-saving and wastewater treatment equipment can encourage industries to invest in water-saving technologies and recycling and to reuse treated wastewater. If water-polluting industries

are charged according to the volume of wastewater they discharge to water bodies, they have an incentive to recycle and reuse water.

11.15 ENVIRONMENTAL WATER DEMANDS

As mentioned previously, in-stream uses are those that utilize water *in situ* as it resides in or moves through the water body itself. In many water bodies, minimum in-stream requirements are specified that may have the highest (inalienable) priority within the water rights system. An environmental water demand (EWD) is the water regime required to sustain the ecological values of an aquatic ecosystem at a low level of risk. If this water requirement is adopted, then it is likely that a water body will:

- Be healthy.
- Look after the needs of animals and plants such as fish, macroinvertebrates, and plants.
- Maintain its biodiversity.

Three steps to identify environmental water requirements in a basin are:

1. Prioritize sub-basins for assessment using information on current water use, river and estuarine health indicators, and water management planning priorities.
2. Consult basin stakeholders and relevant scientific experts to determine important values for each sub-basin in the following broad categories:
   a. Ecosystem
   b. Recreation
   c. Aesthetics
   d. Physical landscape
   e. Consumptive/nonconsumptive use values
3. Assess environmental water requirements using the most appropriate scientific methodology on a catchment-by-catchment basis.

Estimation of environmental water requirements depends on identifying those components of the environment that are to be protected. These components depend upon the amount and timing of water use within the basin. For example, the majority of water use in a catchment may occur in the summer, and this may have an effect on species that have particular requirements in these months (e.g., spawning and fish migration). Other basins may have high water use all year, or dams could be built to alter the flow regime. The rivers may require different forms of assessment to identify higher flows required for such processes as fish spawning and sustaining the ecology of wetlands and estuaries. In heavily regulated rivers where flows can be affected all year by dams or high water use, it is important to understand the timing and duration of high flow events required to maintain the vitality of river systems.

Water chemistry, temperature, nutrient cycling, oxygen availability, and the geomorphic processes that shape river channels and floodplains are often tightly

coupled to streamflow variation. Natural flow regimes are therefore intimately linked to many different aspects of ecological integrity. In the river channel, low flows will concentrate fish and other aquatic organisms, benefiting predators such as larger fish. If low flows are too severe or last for too long due to human influences, large numbers of individuals may perish and jeopardize the local populations of certain species. Thus, rather than trying to prescribe a flow regime that benefits some species all of the time, a better approach is to restore or sustain a flow regime that benefits each species some of the time. The species that are found in each river have endured many trials of adverse flow conditions, exploited many occasions of favorable flow, and have managed to persist in their native rivers over long periods of time. Most environmentalists believe that perpetuation of the natural flow regime is the best approach for conserving the full richness of the biological diversity of a river.

11.16 WATER REUSE

Reused water can be added to the supply system as an additional source of water, thus having a quantitative value, or certain portions of water resources can be reutilized that might otherwise be classified as wastewater, thus having qualitative value. Reused water has been utilized for agricultural and landscaping irrigation, industrial processes, and cooling water while complying with environmental instream flow requirements, groundwater recharge, and direct consumption. Applications of reused water have been increasing as a result of severe droughts and water pollution control regulations (Lund et al., 1995).

Water reuse requires careful control of treatment processes and can be more expensive than the use of freshwater. In evaluating the cost of reuse as a water supply source, the costs of additional treatment, the redistribution system, and operation and maintenance should be considered. The cost of treatment can make it impossible for small communities to reuse water, but large communities may be able to increase their water supply by 50% or more by reusing sewage. Water reuse can be classified as direct and indirect reuse:

- *Indirect reuse*, in which water is taken from a river, lake, or underground aquifer that contains sewage. The practice of discharging sewage to surface waters and withdrawal for reuse allows the processes of natural purification to take place.
- *Direct reuse*, which is the planned and deliberate use of treated wastewater for some beneficial purpose. Direct reuse of reclaimed waters is practiced for several applications without dilution in natural water resources.

Water can be reused for the following purposes:

- *Agricultural irrigation*. The largest user of reused water is agriculture. The original form of sewage treatment was sewage farming, or the disposal of sewage on farmlands. Most sewage farms were replaced 70 to 90 years ago by biological treatment plants, which discharge to the nearest watercourse (Dean and Lund, 1981). Reused water can also be used for urban

irrigation. Irrigation with raw or partially treated sewage can conserve water and fertilize crops economically by capturing nutrients that would normally be wasted. This irrigation method is also an effective way to prevent contamination of nearby waterways with disease organisms such as coliform and other bacteria that sewage contains. The most serious drawback of using sewage for irrigation is its role in transmitting infectious diseases to agricultural workers and the general public. Infections related to *Ascaris* and *Trichuris* worms are commonly associated with wastewater irrigation. Eating uncooked vegetables that have been irrigated with such water could transmit these worms. Wastewater irrigation has also been linked to transmission of enteric diseases such as cholera and typhoid, even in areas where these diseases are not endemic (World Resources, 1998). The advantages of reclaiming wastewater and sewage for irrigation can be summarized as follows:

- Increasing crop yield
- Reduction of pollution of freshwater resources by sewage discharge
- Decreased use of fertilizers
- Soil conservation
- Reduced cost of sewerage and plumbing

- *Urban irrigation*. Urban irrigation includes parks, golf courses, and landscape medians. Limited exposure risk is presented by areas such as golf courses, cemeteries, and highway medians, where public access is restricted and where water is applied only during night hours without airborne drift or surface runoff into public areas. Other urban uses include toilet flushing, fire protection, and construction.
- *Intentional reuse*. Reused water forms part of a potable water supply. The term "intentional" indicates that at least part of the treatment has been given because the wastewater will soon become part of the supply of potable water.
- *Groundwater recharge*. Recharging underground aquifers with treated sewage is one of the most generally accepted forms of water reuse. Water reuse by groundwater recharge is widely believed to entail a lesser degree of risk than other means of reuse in which the recycling connection is more direct. The degree of treatment depends on the type of application to the soil, soil formation and chemistry, depth to groundwater table, dilution available, and residence time to the point of first extraction.
- *Recreation*. Wastewater is sometime used to fill lakes as part of a recreational system. Biological treatment in a series of lagoons usually provides the quality required for recreational purposes.
- *Industrial use*. The quantity of water used in many manufacturing processes and power generation is very large. The treated wastewater is an appropriate source of water for such industries, especially for regions facing water shortages. The water necessary for cooling purposes does not have to be very pure and in some industries can be supplied from treated wastewater.

Various methods have been developed for wastewater treatment, and a brief explanation about these methods and their efficiency in treating sewage from different sources is presented next.

11.17 TREATMENT OPTIONS FOR WASTEWATER REUSE

The main purpose of municipal wastewater treatment is to prevent pollution of the receiving water bodies. The degree of required treatment can be determined based on the beneficial uses of the receiving water bodies. Stream pollution and lake eutrophication resulting from municipal wastewater recharge are particularly troublesome in water reuse projects. Conventional municipal wastewater treatment methods usually consist of the following phases:

- Preliminary treatment
- Primary treatment
- Secondary treatment

Table 11.4 shows a summary of treatment processes required to meet health criteria for wastewater reuse.

11.17.1 Preliminary Process

The preliminary process generally includes pumping, screening, and grit removal. The setup of different units of preliminary treatment depends upon raw-wastewater characteristics, subsequent treatment processes, and the preliminary steps employed. Pretreatment is also necessary when sewage is highly foul smelling because of the presence of hydrogen sulfide, mercaptans, and amines. Chlorine, hydrogen peroxide, chlorine dioxide, and ferric salts can be used to deodorize the sewage. Solids can also be removed in this phase by screening. Flotation is used to remove grease and fats and it can be performed in a pre-aeration tank also used for grit removal. Flocculation may also be practiced on high-strength municipal wastes to provide increased primary removal and prevent excessive loads on the secondary treatment process. Possible arrangements of preliminary treatment units for municipal wastewater treatment are shown in Figure 11.6.

11.17.2 Primary Treatment

In primary treatment, sewage is kept in settling tanks for a specific detention time. Completely mixed activated-sludge processes can be used to treat unsettled raw wastewater; however, this method is restricted to small municipalities because of the costs involved in sludge disposal and operation. Primary sedimentation can remove 30 to 50% of the suspended solids in raw municipal wastewater. A good percentage of biological oxygen demand (BOD) and suspended solids (SS) are removed in this phase. In some cases, lime or ferric salts are added to increase the pH to about 11 in order to precipitate calcium carbonate or calcium phosphate.

TABLE 11.4
Treatment Processes Needed to Meet the Health Criteria for Wastewater Reuse

	Irrigation			Recreational Use			Municipal Use	
	Crops Not for Direct Human Consumption	Crops Eaten Cooked; Fish Culture	Crops Eaten Raw	No Contact	Contact	Industrial Use	Nonpotable	Potable
Health criteria	A+F	B+F/D+F	D+F	B	D+G	C/D	C	E
Primary treatment	Essential	Essential	Essential	Essential	Essential	Essential	Essential	Essential
Secondary treatment		Essential	Essential	Essential	Essential	Essential	Essential	Essential
Sand filtration or equivalent polishing methods		Sometimes necessary	Sometimes necessary		Essential	Sometimes necessary	Essential	May also be essential
Nitrification						Sometimes necessary		Essential
Denitrification						Sometimes necessary		May also be essential
Chemical clarification						Sometimes necessary		May also be essential
Carbon adsorption								May also be essential
Ion exchange or other means of removing ions						Sometimes necessary		May also be essential
Disinfection		Sometimes necessary	Essential	Sometimes necessary	Essential	Sometimes necessary	Essential	Essential[a]

Note: A = freedom from gross solids and significant removal of parasite eggs; B = same as A plus significant removal of bacteria; C = same as A plus more effective removal of bacteria plus removal of viruses; D = not more than 100 coliform organisms per 100 ml in 80% of samples; E = no fecal coliform organisms in 100 ml plus no virus particles in 1000 ml plus no toxic effects on humans plus other drinking water criteria; F = no chemicals that lead to undesirable residues in crops or fish; and G = no chemicals that lead to irritation of mucous membranes or skin.

[a] Free chlorine in one hour.

Source: WHO, Reuse of Effluents: Methods of Wastewater Treatment and Health Safeguards, Technical Report 517, World Health Organization, Geneva, 1973.

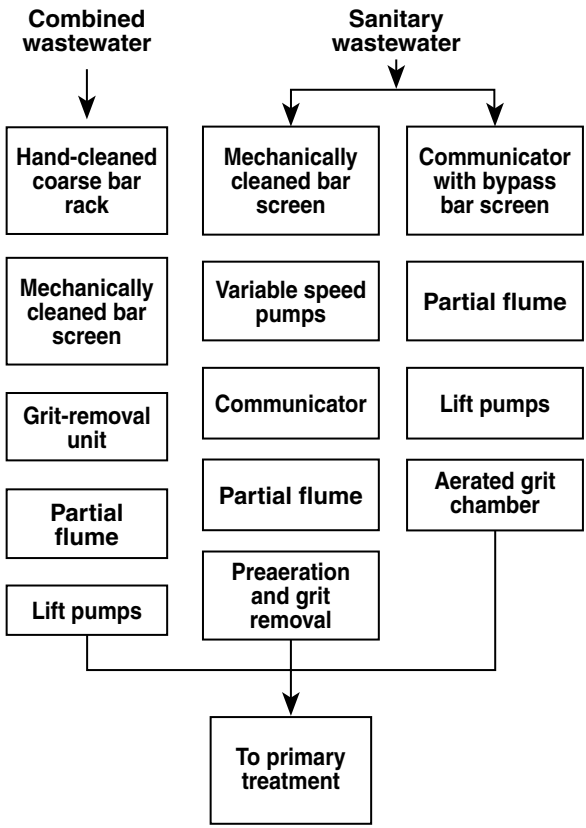


FIGURE 11.6 Three possible arrangements for preliminary treatment units in municipal wastewater processes. (From Viessman, W., Jr. and Hammer, M. J., *Water Supply and Pollution Control*, Harper & Row, New York, 1985.)

11.17.3 Secondary Treatment

Secondary treatment is also known as biological or chemical treatment. The following processes and methods have been widely used for secondary treatment:

- Activated sludge processes
- Trickling filters
- Biological towers

In the activated sludge process, wastewater is continuously fed into an aerated tank, where microorganisms synthesize the organics. The resulting microbial floc (activated sludge) is settled from the aerated mixed liquor under quiescent conditions in a final clarifier and returned to the aeration tank. Advantages of liquid aeration are high BOD removal, ability to treat high-strength wastewater, and adaptability for future use in plant conversion to advanced treatment.

A high degree of operational control is needed in secondary treatment. Shock loads may upset the stability of the biological process, and sustained hydraulic or organic overloading results in process failure. The effluent is made to pass through trickling filters and rotating biological contactors, which provide a media to support the microbial films. Oxygen is supplied from air moving through voids in the media. The residual solids settle out in the clarifier and are removed as stable sludge.

Trickling filters act as contact beds where settled sewage is spread by a rotary distributor over relatively deep (about 2 m) circular beds containing stones 50 to 100 mm in diameter. A typical trickling filter plant consists of primary settling tanks, a trickling filter, final settling tanks, and a recirculation system to maintain a minimum flow through the filter in order to operate the distributor. Trickling filters are more resistant to shock from high organic and hydraulic loads and less expensive to operate than activated sludge plants, but they also cost more to build. Deeper beds up to 12 m deep, *biological towers*, have high-void-ratio plastic media to reduce bed plugging and improve treatment efficiency.

Current models of *rotating biological contactors* (RBCs) are large-diameter polyethylene media with about 40% of the surface area submerged in the waste and rotated slowly on a horizontal shaft in a concrete tank. A biomass film about 1 to 4 mm thick grows on the surface and shears off periodically into the tank below, to be removed by final sedimentation. The process may have oxygen transfer limitations with high-strength wastes; however, for most wastes this method has the lowest power requirement of any competitive biological process (Henry and Heinke, 1989).

11.17.4 Advanced Wastewater Treatment

Advanced wastewater treatment (often called *tertiary treatment*) methods remove more contaminants compared with the above-mentioned conventional methods. Common advanced water and wastewater treatment methods remove:

- Heavy metals
- Organic chemicals
- Suspended solids
- Inorganic salts
- Phosphorus
- Pathogens

They also oxidize ammonia to nitrate. Lime precipitation is one of the suggested methods for removing heavy metals. Applying lime to raw wastewater precipitates phosphates and hardness cations along with organic matter. Suspended solids can also be removed by granular-media filtration. Phosphorus, like the majority of soluble compounds in wastewater, is removed only sparingly by plain sedimentation. About 20 to 40% of phosphorus can be removed through conventional treatment methods. Chemical precipitation using aluminum or iron coagulants is effective in phosphate removal. Chemical/biological phosphorus removal processes also combine chemical precipitation with biological removal of organic matter. Alum or iron

salts are added prior to the primary clarification, directly during the biological process, or prior to final clarification.

Where the ammonia content of the effluent causes pollution of the receiving watercourse, nitrification of wastewater is performed. The process does not remove the nitrogen but converts it to the nitrate form. Nitrate can be reduced to nitrogen gas by heterotrophic bacteria in an anoxic environment. An organic carbon source is required to act as a hydrogen donor and to supply carbon biological synthesis. In biological nitrification and denitrification, unoxidized organic matter can be used as an oxygen acceptor for conversion of nitrate to nitrogen gas.

Pathogens include all categories of microorganisms such as viruses, bacteria, protozoa, and helminths. Disinfection of biologically treated wastewater requires coagulation and granular-media filtration followed by chlorination with an extended contact time. Cysts of protozoa and helminth eggs are resistant to chlorine, and they must be physically removed by effective chemical coagulation and granular-media filtration.

11.18 GRAY WATER: DEFINITION AND REUSE ISSUES

Municipal wastewater has been classified as:

- *Black water* — toilet wastewater (in no-mix toilets, black water can also be classified as brown water or yellow water).
- *Gray water* — household wastewater without black water

Urban sanitation is the first step for controlling water pollution in urban areas. For this purpose, sewage systems were introduced in the middle of the 19th century. They consisted of single-pipe drainage with water-flushed toilets for sanitation. In large cities around the world, surface runoff is also accepted by the single-pipe drainage system that transfers all urban water into a sewage treatment plant. Large amounts of surface runoff containing a variety of pollutants cause inefficiency of most sewage treatment plants.

In a sustainable society, strong coordination between urban and rural components for nutrient and biomass recycling is required. After consumption, the question arises as to how to deal with the resulting urine and feces and include them in the recycling of nutrients and the biomass. One solution is diversion of urine from the conventional sewage pipe drainage outlet. This process of *ecological sanitation* utilizes a dry toilet to divert urine at the source (Winblad et al., 1998). The diverted urine can be then transferred to the nutrient processing plants, where the nitrogen and phosphorus are recovered as chemical fertilizers.

If the urine and feces are collected and used for agriculture, sewage characteristics may change to so-called gray water, which is easier to treat by conventional wastewater treatment processes. The collected gray water can be treated locally and discharged to surface waters in the vicinity of the collection points (Maksimovic and Tejada-Guibert, 2001). The nutrient content of gray water is comparable to waters that by different standards are regarded as “clean” (Gunther, 1996).

A gray-water reuse system should be able to receive the effluent from one or more households in all seasons of the year. Where saturation of garden soils of low permeability occurs in winter rainfall, a facility to divert effluent to sewers or alternative disposal should be present. The system must protect public health and the environment, meet community expectations, and be cost effective (Anda et al., 1996).

Efficient reuse of wastewater and nutrients requires new sanitation designs. Conventional wastewater systems can often handle gray water and municipal wastewater without urine in a reuse-efficient way. Otterpohl and Grottker (1996) suggested the following objectives for designing sanitation systems:

- Sanitation systems must strictly avoid spreading of diseases.
- Nutrient contents of human wastes should be completely recycled to food production (food and water cycles should be kept separate).
- Sanitation systems should have low water/energy consumption and should not promote accumulation of nutrients in ecosystems.
- The use and operation of systems must comply with social and religious rules.
- Costs of installation should be reasonable (an estimate of potential environmental damages must be included).
- Operation should be simple and possible not only for highly specialized experts.

Simple biological treatments can be applied to gray waters. Pretreatment is necessary for separation of solids or grease from kitchen wastewater. Activated sludge systems that rely on settling of sludge can cause problems with the lack of nutrients; therefore, biofilm methods such as trickling filters, biofiltration, rotating disks, or constructed wetlands (bio-sandfilters) are advantageous. Even though gray water does not contain toilet wastewater, it may contain some fecal pathogens from showers and washing machines which makes treatment methods based on sand filtration and membranes preferable. Gray water contains a high amount of household chemicals, thus producing high-quality effluents and reuse without restrictions becomes simpler with the use of household chemicals that can be easily mineralized, not simply degraded (Maksimovic and Tejada-Guibert, 2001).

11.19 CONFLICT ISSUES IN WATER DEMAND MANAGEMENT

Water resources systems generally have multiple objectives, and many of those objectives can conflict. Supplying demand is a main objective of most water resources systems. A common conflict issue in water resources systems planning and operation occurs when the system should supply water to a number of demand points for different purposes with some constraints on water quality, flood control, etc. In many large cities of the world, water demands are raised to a level that

resources are not capable of supplying. Reduction of water demands by water demand management is appropriate in such cases.

In addition to diversity in the objectives of users of water resources, important issues also arise from an institutional point of view. For example, a Department of Agriculture might be interested in supplying irrigation demands and water needed for livestock, while a Department of Industries might be interested in supplying water required for manufacturing processes, cooling systems, etc. In municipalities, responsible agencies are more concerned with increasing the reliability of supplying urban and rural water demands. Thus, the options for water demand management should be selected and applied by the various decision makers who usually come from different agencies and might have different levels of authority to apply their favorite policies. The following example shows how demand management alternatives can be selected by the Nash theorem for resolution of conflict issues.

Example 11.1

Assume city A is located near river X and municipal water demands of this city are supplied from this river. Industrial complex B is also located near this river. Water required for the manufacturing processes and cooling systems of the industries located in this complex is also supplied by a water diversion system implemented downstream of the diversion point for city A. Remaining flow in the river supplies the irrigation demands of agricultural lands located downstream of industrial complex B. The river is capable of supplying total water demands in high flow seasons. For simplicity, the average monthly streamflow in the river and the municipal, agricultural, and industrial water demands are as shown in Table 11.5. As can be seen in this table, about 33% of the water demands cannot be supplied by the river. The following options are proposed for demand management by the responsible agencies:

- Department of Water Supply
  - Public education in the city and vicinities
  - Reduction of unaccounted-for water by leakage control
  - Urban irrigation by extracted water from groundwater aquifer

TABLE 11.5
Average Monthly Streamflow in River X and Water Demands for Example 11.1

	Volume (million m^3)
Streamflow in river X	40
Municipal water demand	20
Industrial water demand	15
Agricultural water demand	25

- Department of Industries
  - Implementation of a system for recycling water for cooling systems in the complex
  - Modification of the manufacturing process
- Department of Agriculture
  - Implementation of drip and pressure irrigation system (for increasing irrigation efficiency)
  - Education and training programs for farmers

Table 11.6 shows the estimated costs and expected percentage of reduction in water demands for each sector. Each of the above-mentioned agencies has proposed a favorite range of investment for demand management based on the costs of each of these options and the benefits associated with supplying water demands with higher reliability. The results of these proposals are summarized in Figures 11.7 to 11.9, where the most favorable range of investment by each organization is indicated by a utility equal to 1. Other investment levels are ranked by lower values of utility. The highest possible level of investment is assigned by each organization for 0 utility. Find the optimal set of demand management options to be able to supply the whole demands of the system by the river.

TABLE 11.6
Estimated Costs and Expected Percentage of Reduction in the Water Demands for Each Sector

Option for Demand Management	Percentage of Reduction in Water Demands of Each Sector	Estimated Cost (monetary units)
Public education in the city and vicinities	1.5	0.8
Reduction of unaccounted-for water by leakage control	15	20.50
Urban irrigation by extracted water from groundwater aquifer	4	3.5
Implementation of a system for recycling water in cooling systems in the complex	45	12
Modification of manufacturing process	12	18
Implementation of drip and pressure irrigation system (for increasing irrigation efficiency)	38	9.5
Education and training programs for farmers	8	0.6

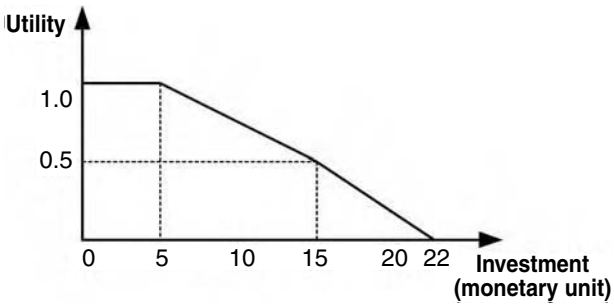


FIGURE 11.7 Utility function of the Department of Water Supply for investing money in municipal water demand management.

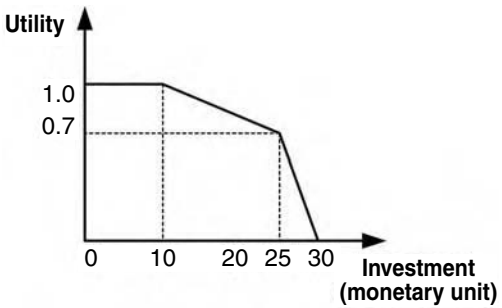


FIGURE 11.8 Utility function of the Department of Industries for investing money in industrial water demand management.

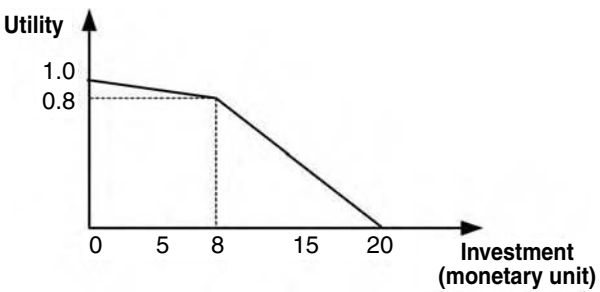


FIGURE 11.9 Utility function of the Department of Agriculture for investing money in agricultural water demand management.

Solution: The nonsymmetric Nash theorem has been used and the optimization problem for finding the optimal set of demand management options is formulated as follows:

$$\text{Maximize} \quad \prod_{i=1}^{3} (f_i - d_i)^{w_i}$$

$$\text{Subject to} \quad d_i \le f_i \le f_i^* \quad \forall i$$

$$\sum_{i=1}^{3} \sum_{j=1}^{NO_i} R_{i,j} \ge 20$$

where w_i is the relative authority of each agency, f_i is the utility function, d_i is the disagreement point, and f_i^* is the ideal point of player (agency) i. In this example, the relative authority of each agency is considered to be equal to one; therefore, the problem is simplified to a symmetric problem. The disagreement point is also set to zero. The above equations show that the total water demand reduction of the system should be more than 20 million m^3, which is the mean monthly deficit in the current situation. In this equation, $R_{i,j}$ is the volume of expected water demand reduction for option j ($j = 1, \ldots, NO_i$), proposed for sector i. NO_i is the total number of water demand management options proposed for sector i. Table 11.7 shows the

TABLE 11.7
Optimal Set of Water Demand Management Options for Example 11.1

Sector	Water Demand Management Option	Reduction in Water Demand (million m^3)	Total Estimated Cost for Each Sector	Utility	Total Water Demand Reduction (million m^3)
Municipal	Public education in the city and vicinity	0.3	21.3	0.041	3.3
	Reduction of unaccounted-for water by leakage control	3			
Industrial	Implementation of a system for recycling water in the cooling systems in the complex	6.75	12	0.96	6.75
Agricultural	Implementation of a drip and pressure irrigation system (for increasing irrigation efficiency)	9.5	10.4	0.66	11.50
	Education and training programs for farmers	2	—	—	—
Total reduction in water demand					21.55

optimal set of water demand management options found by solving the above formulation. As can be seen in this table, the total reduction in water demand is 21.55 million m^3, which is higher than the deficit of 20 million m^3 in the current situation.

11.20 PROBLEMS

11.1 Assume that you want to model the water demand of a city in order to forecast the municipal demands over the next several years. Historical data for water use, population, price, and precipitation for a 10-year time horizon are shown in the following table. Formulate a multiple regression model for estimating the water use data and comment on selecting the independent variables.

Year	Population	Price (monetary unit/m^3)	Annual Precipitation (mm)	Water Use (m^3/year)
1991	21,603	0.25	1015	2,540,513
1992	22,004	0.27	810	2,754,901
1993	23,017	0.29	838	2,722,911
1994	23,701	0.30	965	2,832,270
1995	24,430	0.31	838	2,904,727
1996	25,186	0.32	787	3,004,690
1997	26,001	0.33	720	3,309,927
1998	26,825	0.35	695	3,382,633
1999	27,781	0.31	820	3,375,392
2000	28,576	0.34	800	3,406,259

11.2 Agricultural lands near a river are irrigated by water diverted from the river. The total area of irrigated lands is 400 ha (4,000,000 m^2). Use the following notation to formulate an optimization problem for finding the optimal crop mix. Define the objective function so that as the shortages in supplying irrigation demands are minimized:

$D_{i,j}$ is the irrigation demand of crop i in month j.
$Q_{i,j}$ is the water allocated to crop i in month j.
A_i is the area of the field allocated to crop i.
α is the water supply (irrigation) efficiency
I_j is the river discharge in month j.

11.3 Use the following data to solve Problem 11.2. Assume that the irrigation efficiency is 65%.

	Net Water Demand (mm)											
Crop	**Sept.**	**Oct.**	**Nov.**	**Dec.**	**Jan.**	**Feb.**	**Mar.**	**Apr.**	**May**	**June**	**July**	**Aug.**
Wheat	—	14	5	4	2	1.5	33	142	138	2	—	—
Alfalfa	—	14	5	5	1	2	33	141	97	—	—	—
Cotton	114	61	—	—	—	—	—	24	82	190	225	186

River Discharge (m^3/sec)											
Sept.	**Oct.**	**Nov.**	**Dec.**	**Jan.**	**Feb.**	**Mar.**	**Apr.**	**May**	**Jun.**	**Jul.**	**Aug.**
0.2	0.35	1	1.5	1.7	0.8	0.75	0.6	0.36	0.25	0.2	0.15

11.4 In Problem 11.2, formulate the optimization problem for maximizing the real-time use of river discharge.

11.5 Use the data for Problem 11.3 to solve Problem 11.4.

References

Anda, M., Ho, G., Mathew, K., and Monk, E., Greywater reuse options: areas for further research in Australia, in *Integrated Water Management in Urban Areas*, Niemczynowicz, J., Ed., United Nations Educational, Scientific, and Cultural Organization (UNESCO)-IHP, New York, 1996, pp. 347–356.

Bhatia, R. and Falkenmark, M., Water Resources Policies and the Urban Poor: Innovative Approach and Policy Imperatives, *Water and Sanitation Currents*, United Nations Development Program-World Bank Water and Sanitation Program, New York, 1993.

Blaney, H. F. and Criddle, W. D., Determining water requirements in irrigated areas from climatological data, U.S. Department of Agriculture, Soil Conservation Service, Washington, D.C., 1945.

Bouwer, H., Role of geopurification in future water management, in *Soil and Water Science: Key to Understanding Our Global Environment*, Soil Society of America Special Publication 41, Cincinnati, OH, 1994, p. 7381.

Dean, R. B. and Lund, E., *Water Reuse: Problems and Solutions*, Academic Press, San Diego, CA, 1981.

Falkland, A., Ed., *Hydrology and Water Resources of Small Islands: A Practical Guide*, United Nations Educational, Scientific, and Cultural Organization (UNESCO), New York, 1991.

Gunther, F., Simplifying wastewater treatment by source separation outline for a biologically based greywater purification plant, in *Integrated Water Management in Urban Areas*, Niemczynowicz, J., United Nations Educational, Scientific, and Cultural Organization (UNESCO)-IHP, New York, 1996, pp. 301–314.

Hammer, M. J. and Hammer, M. J., Jr., *Water and Wastewater Technology*, Prentice-Hall, Englewood Cliffs, NJ, 1996.

Hanson, G. and Nilsson, A., Groundwater dams for rural-water supplies in developing countries, *Groundwater*, 24(4), 497–506, 1986.

Henry, J. G. and Heinke, G. W., *Environmental Science and Engineering*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

Lund, J. R., Jenkins, M., and Wilchfort, O., *Integrated Planning and Management for Urban Water Supplies Considering Multiple Uncertainties*, report prepared by the Department of Civil and Environmental Engineering, University of California, Project Number UCLA-WRC-W-813, 1995.

Maksimovic, C. and Tejada-Guibert, J. A., *Frontiers in Urban Water Management*, IWA Publishing, London, 2001.

Mays, L. W., Ed., *Water Resources Handbook*, McGraw-Hill, New York, 1996.

Mays, L. W. and Tung, Y.-K., *Hydrosystems Engineering and Management*, McGraw-Hill, New York, 1992.

Nilsson, A., *Groundwater Dams for Small Scale Water Supply*, Intermediate Technology Publications, London, 1988, 64 pp.

Otterpohl, R. and Grottker, M., Possible sustainable sanitation concepts for urban areas, *Integrated Water Management in Urban Areas*, Niemczynowicz, J., Ed., United Nations Educational, Scientific, and Cultural Organization (UNESCO)-IHP, New York, 1996, pp. 269–278.

Penman, H. L., Natural evaporation from open water, bare soil, and grass, *Proc. Royal Soc.* (London), A, 193, 120–145, 1948.

Solley, W. B., Pierce, R. R., and Perlman, H. A., *Estimated Use of Water Resources in the United States in 1990*, U.S. Geological Survey Circular 1081, Washington, D.C., 1993.

Thomas, H. A., *Improved Methods for National Water Assessment*, Contract WR 1524927, Harvard Water Resources Group, Harvard University, Cambridge, MA, 1981.

Thornthwaite, C. W. and Mather, J. R., The water balance, *Publication in Climatology VIII*, Drexel Institute of Technology, New Jersey, 1955.

Todd, D. K., *Groundwater Hydrology*, John Wiley & Sons, New York, 1980.

United Nations, *The Demand for Water: Procedures and Methodologies for Projecting Water Demands in the Context of Regional and National Planning*, Department of Economic and Social Affairs, United Nations, New York, 1976.

United Nations, *Towards Efficient Water Use in Urban Areas in Asia and the Pacific*, Economic and Social Commission for Asia and the Pacific, United Nations, New York, 1998.

Viessman, W., Jr. and Hammer, M. J., *Water Supply and Pollution Control*, Harper & Row, New York, 1985.

Wegelin-Schuringa, M., *Water Demand Management and the Urban Poor*, IRC International Water and Sanitation Center, at WHO at Kobe Center, 2000.

Winblad, U., Ed., *Ecological Sanitation*, SIDA, Stockholm, 1998.

WHO, *Reuse of Effluents: Methods of Wastewater Treatment and Health Safeguards*, Technical Report 517, World Health Organization, Geneva, 1973.

World Resources Institute, *World Resources: A Guide to the Global Environment*, a joint publication by The World Resources Institute, The United Nations Environment Programme, The United Nations Development Programme, and The World Bank, Oxford University Press, London, 1998.

12 Drought Management

12.1 INTRODUCTION

Population increases and expansion of agricultural and industrial sectors in many parts of the world have resulted in a growing demand for water. The impact of climate change and widespread water contamination has resulted in many irregularities in the supply and demand structure. Drought events have occurred with smaller recurrence intervals with higher peaks and levels of severity. Many limitations have been imposed on water reuse due to quality and environmental restrictions (see Lettenmiare et al., 1996; Aswathanarayana, 2001). All of these factors have led to a need for integrated water resources management that takes into account all of the system components, the interactions between them, and analyses of risk and uncertainty.

In recent years, much attention has been given to labeling droughts based on precipitation data. Precipitation data alone, however, are not sufficient indicators for detecting real droughts based on hydrologic conditions. In fact, the difference between actual demand and user perceptions about water use is the key measure used to evaluate the severity of a drought by hydrologists, meteorologists, water resources engineers, agricultural researchers, economists, and other stakeholders.

Integrated water resources management in a region requires a detailed analysis of normal conditions as well as dry and wet periods. In North America, drought has not been a main focus of water resources studies due to relatively high per capita resources. Recently, though, some attention has been given to droughts in the United States due to water shortages in some areas, such as along the Mexican border. Agreements between the United States and Mexico in regard to water allocation for the Conchos Basin in Mexico and the Rio Grande basin in the United States have been a key factor in promoting drought studies.

This chapter discusses types of drought events, including climatic, hydrologic, and agricultural drought and demonstrates drought analysis and prediction as tools for drought management studies.

12.2 DROUGHT DEFINITION

In general, droughts can be divided into three specific groups: climatic droughts, hydrological droughts, and agricultural droughts. Drought that results from a rainfall deficit is *climatic drought*. For industries and water consumers, drought can occur when the water supply declines; *hydrological drought* is a result of deficit in effective rainfall in a period compared to previous periods. For farmers, drought is a condition that reduces their harvests; *agricultural drought* usually is defined as a reduction of agricultural products and loss of plants. Continuation of rainfall deficits may cause

a permanent shortage in soil moisture in the growth season; therefore, the average shortage in soil moisture over a period of several months for crops that are sensitive to soil moisture and have high water demands can be used as an indicator of drought severity.

A number of definitions of drought events have been suggested by researchers. Yevjevich (1967) stated that if a widespread region is affected by a severe water shortage for a long time, then it is possible to distinguish drought from other events such as water shortage in that region. Dracup et al. (1980) stated that it is not possible to define and analyze drought without paying attention to the nature of water shortage, sequence of time for discretization of drought and truncation levels that are used to separate drought from other phenomena, and regional aspects of the drought.

Based on these definitions, droughts can be divided into two categories: point (localized) or regional. In the point drought analysis, drought in station or small area is investigated. Yevjevich (1967), Griffits (1990), and Beric et al. (1990) have studied point drought. In point investigations, a time series of hydroclimatological factors or a drought indicator is evaluated at a specific point. A time sequence and then a truncation level are assigned to determine drought events in a time series in order to perform the analytical evaluation.

According to Yevjevich (1967), the application of run theory is a fundamental method of analyzing point drought. According to this theory, one could detect a consecutive period of shortage or surplus after choosing a truncation level for existing data by assuming that they are stationary. Then, from detected consecutive shortage periods, probability distributions of drought parameters such as duration and the amount of shortage are evaluated (Guerrero-Salazer and Yevjevich, 1975). Each of these investigators chose a truncation level based on their specific goals. For example, Sen used various probabilities of successful events (wet years) and failure events (shortage). Griffits (1990) used average and Beric et al. (1990) used consecutive time intervals with no rain (15 days) in the growth season as truncation levels.

In the study of regional drought, severe water shortage conditions in contiguous and extended regions are investigated (e.g., Santos, 1983; Sen, 1980). Santos (1983) suggested that, in addition to defining truncation level, a critical area should also be defined. Based on this approach, regional drought occurs when the drought-affected area is greater than the critical area. Many investigators such as Tase and Yevjevich used standardized normal distribution values for drought events with different return periods as a truncation level, assuming that all the specified rain gauges in the study area follow the normal distribution. Santos (1983), who has done intensive studies on the analysis of regional droughts, used 10 to 20% of average regional precipitation as a truncation level.

12.3 DROUGHT PREPAREDNESS/MANAGEMENT

The flowchart of Figure 12.1 describes the types of droughts and how they can be integrated into a drought management/preparedness algorithm (Karamouz et al., 2002). The primary objective of this algorithm is to put climatic, hydrologic, and agricultural droughts into some perspective to understand how they can be related

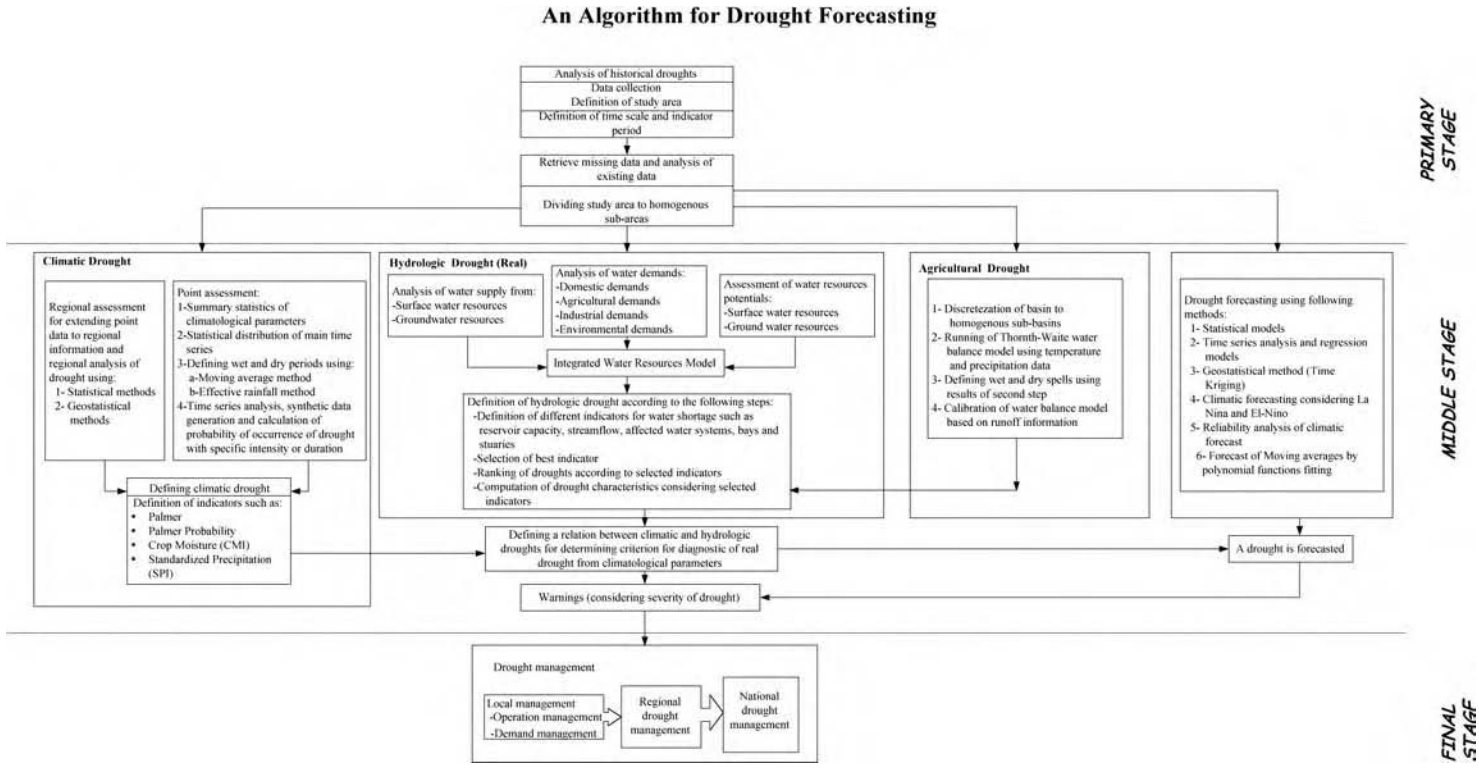


FIGURE 12.1 Algorithm for drought preparedness/management. (From Karamouz, M. et al., *Proc. 2002 EWRI Conference on Water Resources Planning and Management*, Roanoke, VA, 2002.)

to real drought and how their results can be used for development of water resource strategies in a region. The factors affecting hydrologic droughts are so complex and interrelated that they cannot be modeled unless a precise water balance model is developed. Such a water balance model can be used to make a realistic assessment of rainfall, runoff, changes in temperature, and soil moisture conditions. It is impossible to assess the occurrence and continuity of drought events with reasonable certainty. The results of climatic drought, especially on a regional scale, as well as assessment of effective rainfall, moving averages, and determination of drought-affected areas can guide short- and mid-range water and agricultural resources management. Topics such as drought prediction, forecasting, and monitoring can be investigated once the types of droughts, their thresholds, and how they are related are determined.

12.4 CLIMATIC DROUGHT

12.4.1 Point Evaluation

Climatic drought is defined based on rainfall deficit. In point evaluation of climatic drought to detect dry events, the best probabilistic distributions should be chosen to describe the behavior and variation of precipitation at each sub-basin represented by a rain gauge. The concept of effective rainfall and excess deficit and surplus on a monthly scale can be extended to determine the onset, duration, and the severity of drought. The following steps could be taken in point evaluation of climatic drought:

- Fitting statistical distribution
- Analysis of moving averages
- Point evaluation of drought events in monthly scale

12.4.1.1 Fitting Statistical Distribution

Assume that $\{X_i, i=1, \ldots, N\}$ is a discrete point series vector of a drought indicator in N existing rain gauges of the study area. In analysis of climatic drought, the drought indicator is precipitation and the time sequence is usually considered to be annual. To choose the best statistical distribution for the time series of each station, X_i, the following assumptions should be verified:

- Data are independent.
- Precipitation time series are stationary.
- X_i may have different probability distributions.

To choose the best statistical distribution, the classical approach can be used; however, it has many shortcomings. In this approach, first all of the existing data for point (station) i (X_i) should be sorted in ascending order to form a new set ($x_{(1)}$, $x_{(2)}$, $x_{(n)}$) in such a way that $x_{(1)} = \min(X_i)$ and $x_{(n)} = \max(X_i)$. Then, if the data are sufficient (at least 30 values, as recommended by Kottegoda and Rosso, 1998; Mendenhall

et al., 1990), the range between $x_{(1)}$ and $x_{(n)}$ is divided into K intervals and a histogram for point i can be plotted. The best distribution can be chosen by considering the shape of the histogram and comparing it to the shape of different statistical distributions and by using the result of different goodness of fit tests such as the Kolmogorov–Smirnov and chi-square tests. Alternatively, to choose the best distribution the observed data distribution plot can be compared with the expected values from a statistical distribution plot. Quantile–quantile (Q-Q) and probability–probability (P-P) graphs are two methods used for comparing observed values with values from a given distribution. By these methods, the best distributions can be found with more flexibility and accuracy as compared with the classical method.

In general, it is possible to obtain $(n - 1)$ values in the observed sample in such a way as to divide the cumulative frequency of the sample equally into n values. Each of these values is a *quintile*. The objective is to compare the observed value of events with their corresponding values from a statistical distribution function. To prepare a Q-Q graph, the observed values should be ordered ($x_1 \le x_2 \le \ldots \le x_n$) and then these values are plotted against the inversion of the cumulative probability distribution function as:

$$Q = F^{-1}\left(\frac{\left(i - r_{adj}\right)}{\left(n + n_{adj}\right)}\right) \qquad i = 1,2,...,n \tag{12.1}$$

where n_{adj} and r_{adj} are constants and should be ≤ 0.5, and F^{-1} is the inverse of the cumulative distribution function (CDF) for a statistical distribution. This yields a graph of observed values vs. standardized expected values from a specific distribution. The best statistical distribution is the one that falls into a linear regression relationship of high correlation coefficients with the distribution of observed values. n_{adj} and r_{adj} are limiting the p-values for the CDF to fall between 0 and 1.

To prepare a P-P graph, the cumulative distribution function of the observed values is plotted vs. the CDF of a statistical distribution. To plot P-P graphs, after sorting the observed values in ascending order, one axis is the CDF of the observation i vs. a cumulative distribution function of a statistical distribution for the observed values (X_i). If the selected statistical distribution estimates the observed values closely, then in a P-P graph all of the points should be on a 45° line.

To find a good fit for statistical distribution, parameters of a distribution should be calibrated. One of the most sensitive parameters is the *location parameter*, which is a measure of the truncation level (threshold) for a given distribution. For example, in a Weibull distribution, the location parameter has been used in the following density function:

$$\begin{aligned} &f(x) = c/b * \left[(x-\theta)/b\right]^{c-1} * e^{-\left[(x-\theta)/b\right]^c} \\ &0 \le x < \infty \\ &b > 0;\ c > 0;\ \theta > 0 \end{aligned} \tag{12.2}$$

where:

b is the scale parameter.
c is the shape parameter.
θ is the location parameter (threshold).

The feasible range for this distribution is between 0 and infinity. The location parameter (θ) should be less than the smallest observed value. In this case, this value is subtracted from all of the existing data and the Weibull distribution is then fitted.

The scale parameter (b) could also be considered. Variation of this parameter can change the size of the distribution. As can be seen in Eq. (12.2), as b increases, the size of the distribution (the peak value) becomes smaller. In this case, to keep the area under the curve equal to 1, the distribution will spread out more. By changing these parameters, we can obtain a better fit of a Weibull distribution to the observed data. The other types of distributions have similar parameters.

Beta distribution is another suitable type of distribution for predicting the behavior of precipitation data but has not been given much attention in previous drought studies. This distribution may change from a normal distribution to a very skewed distribution and offers the flexibility necessary for drought studies.

The beta distribution has a probability density function of:

$$f(x) = \Gamma(\upsilon+\omega)/\left[\Gamma(\upsilon)\Gamma(\omega)\right] * \left[(x-\theta)^{\upsilon-1} * (\sigma+\theta-x)^{\omega-1}\right] / \sigma^{\upsilon+\omega+1}$$
$$0 < x < 1 \; ; \; \upsilon > 0 \; ; \; \omega > 0 \tag{12.3}$$

where:

Γ is the gamma function.
υ and ω are the shape parameters.
θ is the threshold (location) parameter.
σ is the scale parameter.

Note that the shape of the beta distribution depends on the values of the parameters ω and υ. The standardized beta distribution has a valid range from 0 to 1. Smaller thresholds can be considered by standardizing the variable as $(x - \theta)/\sigma$. Note that the location parameter must be less than the smallest observed value, and the $(\sigma + \theta)$ must be greater than the largest observed value. In general, if the points in the Q-Q plot form a straight line, then the respective family of distributions (a beta distribution with υ and ω parameters, in this case) provides a good fit to the data; thus, the intercept and slope of the fitted line can be interpreted as graphical estimates of the threshold (θ) and scale (σ) parameters, respectively.

12.4.1.2 Analysis of Moving Averages

In the previous section, the goal was to find the best distribution and the best criteria to separate dry events from the others. In the moving average method, the cumulative

TABLE 12.1
Average Precipitation Data for 10 Years for Two Regions (Example 12.1)

Year	Average Precipitation of Region A (mm)	Average Precipitation of Region B (mm)
1	640	230
2	456	127
3	236	261
4	456	301
5	235	261
6	601	322
7	532	400
8	731	223
9	256	128
Current year	541	112

impact of precipitation deficits and surpluses in previous years are considered. By using this method, the drought investigation is diverted from climatic drought to hydrological (real) drought. The n-year moving average, X'_j in the jth year, can be calculated from the following expression:

$$X'_j = \frac{1}{n}\sum_{i=0}^{n-1} X_{j+i} \qquad n = 3,5,9,... \tag{12.4}$$

Usually, 3-, 5-, and 10-year moving averages are investigated to recognize drought spells and to predict their trend.

Example 12.1

Use the 3-year moving averages to compare the two series of annual average precipitation for two different regions presented in Table 12.1. Which of these regions is facing a severe drought in the current year? If the long-term average precipitation in regions A and B is estimated to be 400 and 250 mm, respectively, discuss the drought periods and their durations.

Solution: Table 12.2 shows the estimated 3-year moving averages for regions A and B. Figures 12.2 and 12.3 show the 3-year moving average and long-term average of precipitation for the two regions. Region A has experienced one drought event of a 2-year duration. The current year in that region is experiencing a wet spell. Region B was in the middle of a drought period for the first years of the record, and the current year seems to be the first year of a drought event.

TABLE 12.2
3-Year Moving Average for Two Regions (Example 12.1)

Year	3-Year Moving Average for Region A (mm)	3-Year Moving Average for Region B (mm)
1	—	—
2	—	—
3	444.00	206.00
4	382.67	229.67
5	309.00	274.33
6	430.67	294.67
7	456.00	327.67
8	621.33	315.00
9	506.33	250.33
Current year	509.33	154.33

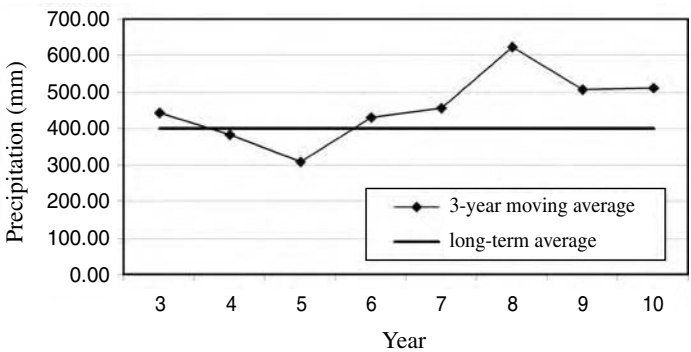


FIGURE 12.2 Long-term average annual precipitation and 3-year moving average for region A in Example 12.1.

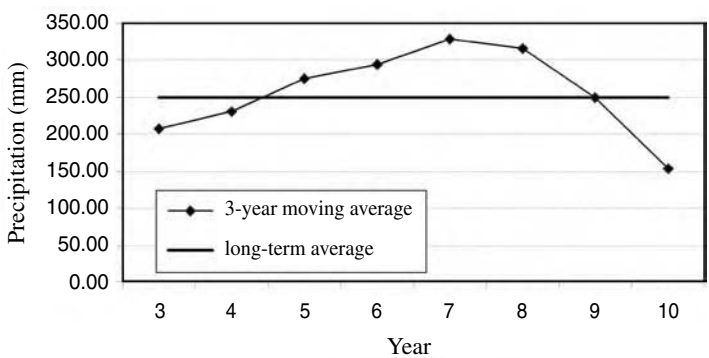


FIGURE 12.3 Long-term average annual precipitation and 3-year moving average for region B in Example 12.1.

12.4.1.3 Point Evaluation of Drought Events on a Monthly Scale

In this method, which was initially proposed by Herbst et al. (1966), it is assumed that farming in any region is adapted to the prevailing climatic pattern so that maximum advantage is taken of the months of high average rainfall (Chang, 1990). Due to variations in mean monthly rainfall, a seasonal drought of a certain severity and duration is often a normal feature of that particular climate and as such should not be included in the assessment of droughts with severe consequences. In the following sections, it is assumed that the benefit to vegetation continues for some time after excessive rain due to the storage of moisture in the soil and, conversely, that the recovery of vegetation after a drought is not immediate and the impact persists for some time.

12.4.1.3.1 Effective Rainfall and Rainfall Difference Series

In the literature, effective or excess rainfall refers to that portion of precipitation that is converted to runoff, but in drought studies the definition of *effective rainfall* is different and can be calculated as follows:

$$\begin{aligned} &ER_{i,n} = M_{i,n} + W_i\left(ER_{i-1,n} - \overline{M}_{i-1}\right) \qquad (n = 1,\ \ldots,\ NY)\ ,\ (i = 1,\ldots,12) \\ &ER_{1,1} = M_1 \end{aligned} \tag{12.5}$$

$$W_i = C\left(1 + \frac{\overline{M}_i}{\frac{1}{12} \times MAR}\right) \tag{12.6}$$

where:

MAR is the mean annual rainfall.
C is a constant.
$ER_{i,n}$ is the effective rainfall in month i of year n.
$M_{i,n}$ is the rainfall in month i of year n.
$\overline{M}_i$ is the long-term monthly mean for month i (i = 1, …, 12).
W_i is the weighting factor for month i.
NY is the number of years of the record.

C is a constant that can be obtained by trial and error in order to better simulate the observed wet and dry periods. In the previous studies a value of about 0.1 is selected for C. The weighting factors obtained from this equation show the effect of rainfall at each month on the drought situation of the next months. An average deficit ($\overline{DF}_i$) for month i is defined as:

$$\overline{RD}_i = \frac{\sum_{n=1}^{NY} DELTA_{i,n}}{NY} \qquad (i = 1,\ldots,12) \tag{12.7}$$

$$DELTA_{i,n} = \begin{cases} \overline{M}_i - ER_{i,n} & ER_{i,n} < \overline{M}_i \\ 0 & ER_{i,n} \geq \overline{M}_i \end{cases} \qquad (n = 1, \ldots, NY) \qquad (12.8)$$

The carryover from month to month is defined as the effect of rainfall deficit or surplus from the previous month:

$$ERS_{i,n} = \begin{cases} ER_{i,n} - \overline{M}_i & \text{if} \quad ER_{i,n} \geq \overline{M}_i \\ 0 & \text{otherwise} \end{cases} \qquad (12.9)$$

$$ERD_{i,n} = \overline{M}_i - ER_{i,n} - \overline{RD}_i \qquad (12.10)$$

$$RD_{i,n} = Min\left[(M_i - ER_{i,n}), \overline{RD}_i\right] \qquad (n = 1, \ldots, NY) \qquad (12.11)$$

where:

$ERS_{i,n}$ is the excess rainfall surplus in month i of year n.
$ERD_{i,n}$ is the excess rainfall deficit in month i of year n.
$RD_{i,n}$ is the rainfall deficit in month i of year n.
$\overline{RD}_i$ is the long-term monthly mean for rainfall deficit in month i.

The carryover from month to month is determined by subtracting the mean rainfall for a particular month from the effective rainfall for the same month so that a deficit or surplus rainfall for that month is obtained. The deficit and surplus rainfall is multiplied by a weighting factor for the next month and the product, whether negative or positive, is added algebraically to the rainfall figures of that month. This sum is the effective rainfall. The weighting factor is used to calculate the carryover effect of a particular month and shows the relative effect of the deviation from the mean in a month as it contributes to the soil moisture of the following month.

For these equations, the carryover is assumed to be 0 for the first month of the record. The process is continued to obtain the effective monthly rainfall for the entire period of the record. *Excessive rainfall deficit* (ERD) is of particular importance to drought studies. This measure indicates how large the deficit of a system could tolerate beyond the normal amount of deficits of the system naturally and over the coarse of experienced historical data. This parameter is the main feature of calculating drought severity.

Example 12.2

The monthly rainfall data for a period of 24 months and the long-term monthly average rainfall are presented in Table 12.3. The mean annual rainfall (*MAR*) is estimated as 535.2 mm from long-term data. Assume *C* (the constant in Eq. (12.6))

TABLE 12.3
Rainfall Data for Example 12.2

Month	Rainfall (mm)	Mean Monthly Rainfall (mm)
1	0.99	1.25
2	3.99	1.13
3	1.59	0.67
4	19.59	10.43
5	2.45	57.51
6	21.06	77.7
7	72.31	76.75
8	75.21	80.91
9	116.42	106.08
10	34.81	75.22
11	14.62	43.37
12	0.00	4.17
13	1.02	1.25
14	0.86	1.13
15	0	0.67
16	0.7	10.43
17	67.4	57.51
18	51.7	77.7
19	45.4	76.75
20	81.2	80.91
21	38.1	106.08
22	57.1	75.22
23	2.3	43.37
24	0.1	4.17

is 0.1. Calculate the effective rainfall, rainfall deficit, and excess rainfall deficit for the region.

Solution: Equations (12.5) and (12.6) are used for estimating the effective rainfall. Based on Eq. (12.5), the effective rainfall in the first month is considered to be equal to the average rainfall in that month, which is 1.25 mm. For the second month, the weighting factor can be estimated as follows:

$$W_2 = 0.1 \times \left(1 + \frac{1.13}{\frac{1}{12} \times 535.2} \right) = 0.103$$

Then, the effective rainfall for this month can be estimated using Eq. (12.5):

$$ER_{2,1} = 3.99 + 0.103 \times (0.99 - 1.25) = 3.96 \text{ mm}$$

Equations (12.7) and (12.8) are used for estimating the rainfall deficit. For example, for the first month, it can be written:

$$ER_{1,1} = M_1 < \overline{M}_1 \Rightarrow DELTA_{1,1} = 1.25 - 0.99 = 0.26 \text{ mm}$$

$$ER_{1,2} > \overline{M}_1 \Rightarrow DELTA_{1,2} = 1.25 - 0.18 = 1.07 \text{ mm}$$

Therefore,

$$\overline{RD}_1 = \frac{\sum_{n=1}^{2} DELTA_{i,n}}{2} = \frac{0.26 + 1.07}{2} = 0.665 \text{ mm}$$

Using Eq. (12.11), it can be written that:

$$RD_{1,1} = \text{Min}\left[(0.99 - 0.99), 0.665\right] = 0.0 \text{ mm}$$

The excess rainfall deficit for the first month is estimated using Eq. (12.10) as follows:

$$ERD_{1,1} = 1.25 - 0.99 - 0.665 = 0.405 \text{ mm}$$

Table 12.4 shows the values of effective rainfall, rainfall deficit, and excess rainfall deficit estimated for the 15 months of data presented in Table 12.3.

12.4.1.3.2 Onset and Termination of Drought

In a test to signify the onset of a drought, it can be assumed that a single month could constitute a drought only if no rain occurred in the month of highest mean rainfall. The test is based on a comparison of the sum of the negative differences (deficits) from the point in time the test begins, with a sliding scale of 12 values calculated by linear interpolation between the maximum value of the mean monthly rainfall (*MMMR*) and the mean annual deficit (*MAD*). A monthly increment x is thus obtained from the following equation:

$$x = \frac{MAD - MMMR}{11} \tag{12.12}$$

where $MMMR = \text{Max}[\overline{M}_i]$, and *MAD* is equal to the ratio of total deficits in the historical period to the number of the years in which deficits have occurred.

The sliding scale is a line connecting the *MAD* to the *MMMR*. The first value on the sliding scale is equal to the *MMMR* and is the maximum deficit that can occur in a single month. The second value on the sliding scale is obtained by adding x to the *MMMR*; the third, by adding $2x$; and so on up to $MMMR + 11x$, which is equivalent to the *MAD*. To start the test, it can be assumed that no drought prevailed prior to the start of the available rainfall record. If the first month of available data

TABLE 12.4
Effective Rainfall, Rainfall Deficit, and Excess Rainfall Deficit Estimates for Example 12.2

Month	Effective Rainfall (mm)	Rainfall Deficit (mm)	Excess Rainfall Deficit (mm)	Effective Rainfall — Average Monthly Rainfall (mm)
1	0.99	0.00	0.13	–0.26
2	3.96	0.03	–2.18	2.83
3	1.88	–0.29	–0.85	1.21
4	19.74	–0.15	–8.13	9.31
5	4.58	–2.13	52.87	–52.93
6	6.55	6.21	69.10	–71.15
7	52.95	12.93	14.08	–23.80
8	68.51	6.70	3.72	–12.40
9	112.23	3.84	–9.82	6.15
10	36.46	–1.65	29.98	–38.76
11	6.98	7.50	28.96	–36.39
12	–3.98	3.98	3.51	–8.15
13	0.18	0.27	0.80	–1.07
14	0.75	–1.34	1.71	–0.38
15	–0.04	–0.43	1.14	–0.71
16	0.61	–2.20	12.02	–9.82
17	65.15	2.25	–17.05	7.64
18	53.80	–2.10	–11.81	–23.90
19	38.90	6.50	10.54	–37.85
20	70.55	10.65	–0.51	–10.36
21	34.60	3.50	56.69	–71.48
22	37.90	19.20	–0.36	–37.32
23	–5.06	7.36	9.03	–48.43
24	–5.20	5.30	0.91	–9.37
Sum.	—	85.93	—	—

with negative difference $(M_i - \overline{M}_i)$ could be considered as the start of a potential drought, then a comparison is made between the absolute value of this deficit and the first value of the sliding scale. If the absolute value of $(M_i - \overline{M}_i)$ exceeds the sliding scale, then a drought is deemed to have started. Likewise, if the difference of the next month is checked and if it is negative, then it is added to the negative difference of the first month and compared with the second value on the sliding scale. If this criterion is exceeded by the absolute value of the two deficits combined, a drought is deemed to have started from the first month and the process continues until the sliding scale reaches the *MAD*.

In the test of drought termination, the test is applied to the period following the first month with a positive difference (surplus) occurring after the start of a drought. The test for termination of a drought consists of either a continuous period of slightly above-average rainfall or a few successive months of abnormally high rainfall, indicating drought termination. The duration of the drought is the period between the onset and termination of the drought. One way to determine drought termination

is to look at successive surplus periods to see how they compare with the maximum of 3-, 4-, …, 12-month successive averages. If the successive surpluses exceed the maximum successive averages, then the drought could be considered terminated. After determining a drought period, its severity is obtained from the following equation (for more details, see Herbst et al., 1966):

$$S_d = \frac{\sum_j (ERD_j) * D_d}{\sum_d \overline{RD}_d} \tag{12.13}$$

where:

S_d is the severity of drought d.
d is the index of drought.
j is all the months of excess rainfall deficit in drought d.
D is the drought duration (month).
$\overline{ERD}_j$ is the excess rainfall deficit observed in month j.
$\overline{RD}_d$ is the average deficit in each month of period d.

Example 12.3

Determine the onset, duration, and severity of drought for the data presented in Example 12.2. Consider that the mean annual deficit is 150 mm and the maximum values of the n-month average rainfall are shown in Table 12.5.

Solution: As shown in Table 12.3, the maximum value of the mean monthly rainfall (*MMMR*) is 106.08 mm. The sliding scale is then estimated using Eq. (12.12) as follows:

$$x = \frac{150 - 106.08}{11} = 3.99$$

TABLE 12.5
Maximum *n*-Month Average of the Rainfall Series (mm) for Example 12.3

Maximum 2-month average	175.0 mm
Maximum 3-month average	243.1 mm
Maximum 4-month average	265.3 mm
Maximum 5-month average	275.3 mm
Maximum 6-month average	283.9 mm
Maximum 7-month average	298.4 mm
Maximum 8-month average	310.5 mm
Maximum 9-month average	345.6 mm
Maximum 10-month average	366.3 mm
Maximum 11-month average	380.5 mm
Maximum 12-month average	410.1 mm

TABLE 12.6
Estimated Values for the Sliding Scale for Example 12.3

Month	Sliding Scale Value
1	106.08
2	110.07
3	114.06
4	118.05
5	122.04
6	126.03
7	130.02
8	134.01
9	138.00
10	141.98
11	145.97
12	150.00

TABLE 12.7
Cumulative Differences and Sliding Scales for Example 12.3

Month	Sliding Scales	Cumulative Difference
1	106.08	–0.26
2	110.07	2.57
3	114.06	3.78
4	118.05	13.09
5	122.04	–39.84
6	126.03	–110.99
7	130.02	–134.79
8	134.01	–147.19
9	138	–141.04
10	141.98	–179.80
11	145.97	–216.19
12	150.00	–224.34

Table 12.6 shows the values estimated for the sliding scale for different months. To test the onset of droughts, the differences between effective rainfall and the long-term average of rainfall in each month are estimated, which is shown in Table 12.4.

As it can be seen in Table 12.4, the first negative difference occurs in month 1; therefore, it could be considered as the onset of a drought event. The cumulative summations of the differences after the first negative value are compared with the sliding scales (x) in Table 12.7. As shown in this table, the absolute values of the difference series become greater than the sliding scale series after month 7. Now it

TABLE 12.8
Results of Drought Termination Analysis for Example 12.3

Month (1)	Rainfall (2)	Differences = Effective Rainfall – Monthly Mean Rainfall (3)	Cumulative Rainfall After the First Positive Difference (4)	Maximum *n*-Month Averages (5)
1	0.99	–0.26	—	—
2	3.99	2.83	—	—
3	1.59	1.21	—	—
4	19.59	9.31	—	—
5	2.45	–52.93	—	—
6	21.06	–71.15	—	—
7	72.31	–23.80	—	—
8	75.21	–12.40	—	—
9	116.42	6.15	116.42	—
10	34.81	–38.76	151.23	175.0
11	14.62	–36.39	165.85	243.1
12	0.00	–8.15	165.85	265.3
13	1.02	–1.07	166.87	275.3
14	0.86	–0.38	167.73	283.9
15	0	–0.71	167.73	298.4
16	0.7	–9.82	168.43	310.5
17	67.4	7.64	235.83	345.6
18	51.7	–23.90	287.53	366.3
19	45.4	–37.85	332.93	380.5
20	81.2	–10.36	414.13	410.1
21	38.1	–71.48	—	—
22	57.1	–37.32	—	—
23	2.3	–48.43	—	—
24	0.1	–9.37	—	—

can be said that a drought has started from month 1. If this test is not passed after considering 11 sliding scales, the test is performed again after observing another negative difference.

The test of termination of a drought starts after the first positive difference between effective rainfall and monthly mean rainfall is observed during a drought. In this example, the application of the test starting from first positive value shows that the drought has not been terminated yet. In the second iteration, the cumulative summation of rainfall (column 4) is compared with the maximum *n*-month averages (column 5), which shows that the values in column 4 become greater than the values in column 5 after the 20th row. Therefore, it can be said that the drought has been terminated and 20 months is considered as the duration of drought.

To calculate the severity of this drought, Eq. (12.13) is used as follows:

$$\sum_j \left(ERD_j\right) = 212.31$$

$$D_d = 20.0$$

$$\sum_d \overline{RD}_d = 233.7$$

$$S_d = \frac{212.31 \times 20}{233.7} = 18.1$$

12.4.2 Regional Analysis of Climatic Drought

The main objective for regional analysis is to find a suitable criterion for defining dry events on a regional scale. Three methods can be used for this analysis: weighting, median of ratios, and combined methods.

12.4.2.1 Weighting Method

In this method, the weights of each station are calculated by applying the Thiessen polygon network and the long-term average regional precipitation dataset is obtained for the entire region. Then, the best statistical distribution is chosen for the long-term average precipitation dataset ($\overline{P}$) and based on that, the ratios of average regional precipitation to long-term average precipitation for various probabilities (different return periods) are calculated. These ratios are applied as regional criteria to designate the precipitation datasets of each rain gauge of the region as wet or dry events.

Example 12.4

The historical annual precipitation data for a basin are given in Table 12.9. Table 12.10 shows the annual rainfall in drought events with different return periods. If a 10-year climatic drought occurs in the region, determine which sub-basins face a drought with a 10-year or more return period.

Solution: In order to find the annual rainfall associated with a 10-year drought event, the data given in Table 12.9 are sorted in ascending order (see column 3 of Table 12.9). Then, the probability for each data is estimated using the following formula:

$$p = \frac{m}{n+1}$$

where m is the rank of the sorted data and n is the total number of data.

TABLE 12.9
Annual Average Precipitation for a Basin (Example 12.4)

No. of Years (1)	Precipitation (mm) (2)	Sorted Precipitation (mm) (3)	Probability (4)	Return Period (yr) (5)
1	618.23	101.47	0.05	19.00
2	563.54	284.2	0.11	9.50
3	546.6	307.44	0.16	6.33
4	501.51	345.81	0.21	4.75
5	480.41	350.1	0.26	3.80
6	451.88	368.51	0.32	3.17
7	418.34	369.99	0.37	2.71
8	403.34	403.34	0.42	2.38
9	369.99	418.34	0.47	2.11
10	368.51	451.88	0.53	1.90
11	350.1	480.41	0.58	1.73
12	345.81	501.51	0.63	1.58
13	307.44	546.6	0.68	1.46
14	284.2	563.54	0.74	1.36
15	776.02	617.04	0.79	1.27
16	1001.96	618.23	0.84	1.19
17	101.47	776.02	0.89	1.12
18	617.04	1001.96	0.95	1.06

TABLE 12.10
Estimated Regional Precipitation Data for Drought Events (Example 12.4)

Sub-Basin	Area (%)	Drought Event Precipitation (mm)				
		2-Year	5-Year	10-Year	25-Year	50-Year
A	20	380	280	180	100	70
B	30	600	540	510	490	350
C	15	550	520	480	320	250
D	15	280	250	200	180	70
E	20	400	350	310	280	190

TABLE 12.11
Recorded Precipitation Data for Stations in Example 12.5

	Stations Representing Sub-Basins				
	A	B	C	D	E
Annual precipitation (mm)	380	520	320	250	410

As can be seen in Table 12.9, annual precipitation of 284 mm has a 10-year return period. Sub-basins A and D are facing droughts with 10-year and more return periods.

Example 12.5

In a specific year, the data shown in Table 12.11 have been recorded in the region presented in Example 12.4. What is the appropriate return period of the regional drought?

Solution: The average precipitation over the region can be estimated using the Thiessen polygon network as follows:

Average rainfall =

$$\frac{(380\times 20)+(520\times 30)+(320\times 15)+(250\times 15)+(410\times 20)}{100}=399.5$$

Considering the data presented in Table 12.9, the regional drought is estimated to have a 2.4-year (between 2 and 3 years) return period.

12.4.2.2 Median of Ratios Method

The objective is to obtain a regional truncation level; therefore, the medians of the calculated P/P_{mean} ratios are determined for various return periods for point evaluation of the different stations in the region:

$$\left(\frac{P}{P_{mean}}\right)_T = Median\left\{\left(\frac{P}{P_{mean}}\right)_i^T; i=1,...,N\right\}; \qquad T=2,5,10,... \qquad (12.14)$$

where:

$(P/P_{mean})_T$ is the regional criterion of precipitation ratio to long-term average precipitation for return period T.
$(P/P_{mean})_i^T$ is the ratio of precipitation to long-term average precipitation in rain gauge i for return period T.
N is the number of total rain gauges in the study region.

12.4.2.3 Combined Method

This method is a combination of the previous two methods. In this method, the estimated values for $(P/P_{mean})_T$ in the point evaluations for each station will be weighted by using their relative area in the Theissen network. Therefore, the third regional criterion can be calculated as follows:

$$\left(\frac{P}{P_{mean}}\right)_T = \sum_{i=1}^{N}\alpha_i \times \left(\frac{P}{P_{mean}}\right)_i^T \quad ; \qquad T=2,5,10,... \qquad (12.15)$$

In addition to truncation level, other important factors should be considered in regional drought analysis. Santos (1983) suggested that, in addition to defining critical and truncation levels, a critical area should also be defined. Regional drought is said to occur when the drought-affected area is larger than the critical area. The critical area can be considered to be equal to a percentage of the total area, depending upon the severity of the decisions imposed by using this threshold. For example, if the threshold is selected as 50% of the area of a region, then the corresponding return period that results in drought for an affected area equal to or greater than 50% is considered to be the specific return period for that year.

The other factor that should be considered in drought regionalization is the range of variation for point averages in the entire region. For this purpose, dimensionless precipitation has been used, which is a ratio of precipitation to average precipitation. Rossi et al. (1992) suggested the use of dimensionless precipitation. The drought-affected area in the current period is defined and analyzed by applying the observed data, as follows:

- The total observed data for all of the rain gauges are changed to be dimensionless.
- A regional isohyetal map is plotted over the entire region.
- Based on regional criteria definitions, drought-affected areas are estimated by summation of all of the drought-affected subareas over different time periods as follows:

$$A_D^T(t) = \sum \beta_i \cdot a_i(t)$$
$$\beta_i = \begin{cases} 1 & \text{if} \quad P_i(t) \le P_{iT} \\ 0 & \text{otherwise} \end{cases} \tag{12.16}$$

where:

$A_D^T(t)$ is the drought-affected area with return period T during time t.
$a_i(t)$ is the subarea represented by station i during time t.
$P_i(t)$ is the precipitation in station i in a given time period t.
P_{iT} is the precipitation at station i with a return period T obtained from the best-fitted distribution for that station.

Figure 12.4 shows an isohyetal map of P/P_{mean} ratios for a 25-year return period (drawn for a river basin in the central part of Iran) (Karamouz et al., 2002). This map can be plotted using the spatial kriging method. Similar maps can be generated for different return periods. According to these maps, if the region is affected by a drought with return period T, one can identify parts of the region that are in the most critical situation.

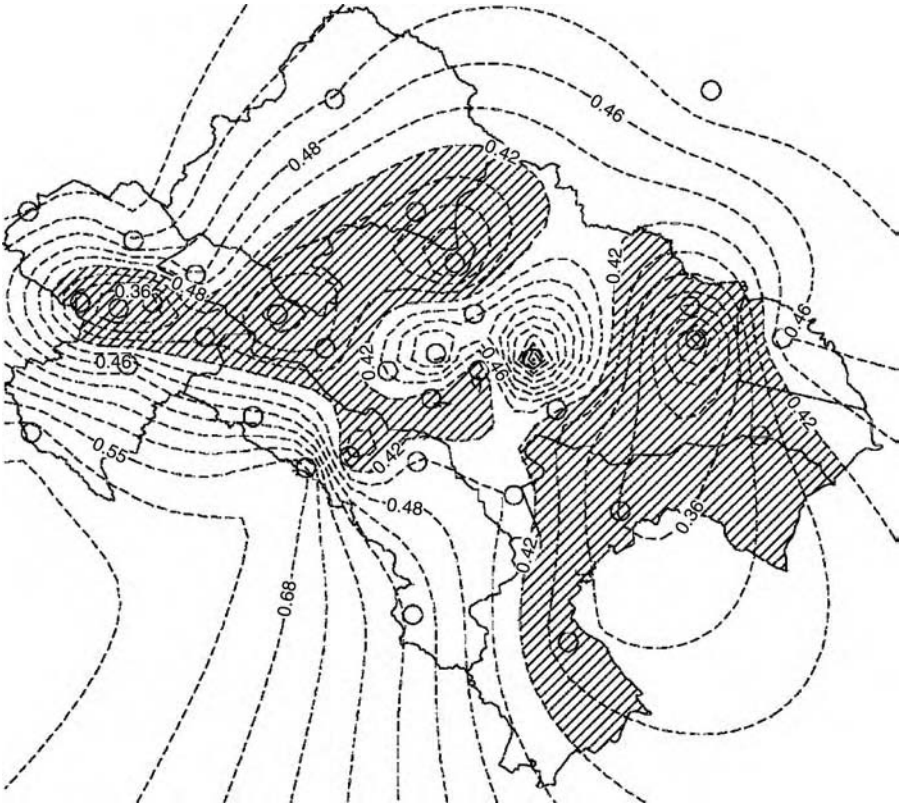


FIGURE 12.4 Isohyetal map for P/P_{mean} ratios for 25-year return period for a river basin in the central part of Iran (the hatched sections show the drought affected areas in the region). (From Karamouz, M. et al., *Proc. 2002 EWRI Conference on Water Resources Planning and Management*, Roanoke, VA, 2002. With permission.)

12.5 PREDICTION OF DROUGHT TRENDS

One of the most important considerations for decision makers, operators, and managers in drought condition is to predict the drought trends and how long a drought will last based on previous climatic data. Among the existing methods, the Poisson statistical method and kriging geostatistical method have been widely used. The time between termination of drought and the start of the next drought event, known as the expected waiting time, follows a Poisson distribution. The kriging geostatistical method is used to predict inflow to a reservoir during critical water-shortage periods in drought-affected regions. If the predicted inflow is not enough to meet the estimated demand, then a drought has begun. If the time between the existing drought and the next expected drought event (start time forecast for drought) is shorter than the remaining water shortage period, then the operation strategies should be changed from normal to water rationing. This method has been suggested by Rouhani and Cargile (1989).

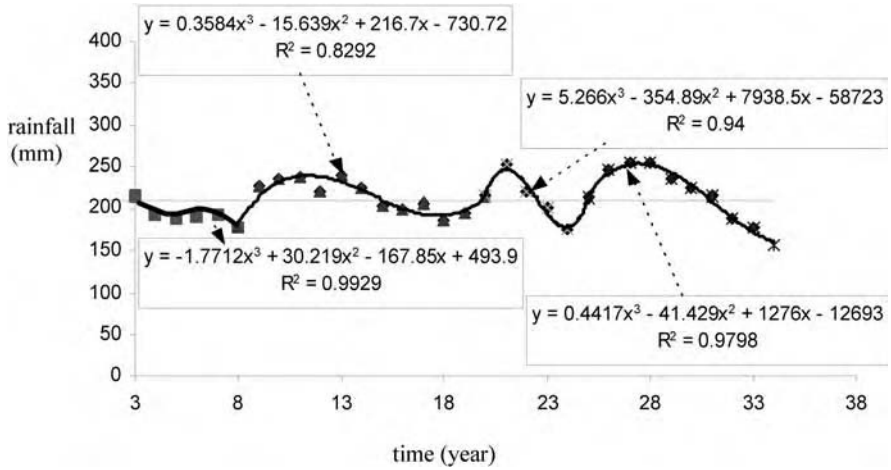


FIGURE 12.5 Three-year moving average of precipitation the Zayandeh-rud River basin in the central part of Iran and polynomials fitted to the dry and wet spells. (From Karamouz, M. et al., *Proc. 2002 EWRI Conference on Water Resources Planning and Management*, Roanoke, VA, 2002. With permission.)

A third method is based on long-term regional average precipitation data and its moving average has also been investigated by Karamouz et al. (2002). In this method, the spells can be approximated by a polynomial function as follows:

$$Y_{ki} = a_{0k} + a_{1k} \times t_{ik} + a_{2k} \times t_{ik}{}^2 + \ldots + a_{nk} \times t_{ik}^n \quad ; \; k = 1,2,..., \; K; \; i = 1,2,..., T_k' \quad (12.17)$$

where:

Y_{ki} is the moving average value in the kth wet and dry spells during sequence i.
a_{jk} is the jth polynomial coefficient in the kth wet and dry spell.
t_{ik} is the time sequence for the kth wet and dry spell.
k is the number of existing wet and dry spells in the observed dataset.
T_k' is the wet and dry spell time period.

If a region is in the middle of a dry spell, a polynomial can be fitted to predict the duration of the drought. Figure 12.5 shows the 3-year moving average and fitted polynomial for the dry and wet spells for a river basin in the central part of Iran (Karamouz et al., 2002).

12.6 HYDROLOGIC DROUGHT: DEFINITION AND BACKGROUND

Hydrologic drought investigation consists of the analysis of water supply and demand as well as socioeconomic factors. In general, *hydrologic drought* occurs when the water resources in a region cannot supply water demands. To investigate the available

and potential water supply of a system, surface reservoirs, groundwater, and inflows to the system must be considered. As far as water demands, different demands of the system and their trends and variations, especially during drought periods, should be studied. Natural and uncontrolled events include streamflows; others are induced by human intervention, such as the operation of reservoirs. Because of the natural uncertainty in streamflow and its relation with other parameters such as precipitation, the most important part of a hydrologic drought study is streamflow investigation.

For hydrologic drought, most of the literature has concentrated on analysis of historical and generated streamflow data, including estimation of the pdf (probability density function) of the drought characteristics. These efforts can be classified into three categories:

- *Experimental investigation of data* attempts to describe drought characteristics based on observed series such as streamflows and demands. Based on this method, Chang (1990) defined the expected time between drought events using a Poisson distribution, and then he investigated the variation in streamflow behavior during a drought using daily streamflow data. Dracup et al. (1980) studied stationary and stochastic characteristics of data and the correlation between various drought characteristics in the historical data. They suggested that the results based on observed data may be insufficient because of a short range of recorded data. Ben-Zvi (1987) considered the volume of streamflows to be an indicator of drought. He defined average data as the truncation level of water shortages during small droughts and standard deviation as the truncation for severe droughts and then studied droughts of different severity. Yevjevich (1967) proposed that the difference between water supply and demand could be used as an indicator of hydrologic drought. He defined sequential periods (runs) as drought spells in series and emphasized that run theory is the best method of studying hydrologic droughts.
- *Investigation of generated data* applies the previous methodology to generated data. Different methods may be used to generate data. Zucchini and Adamson (1988) used the bootstrap method to evaluate the risk of shortage of annual inflow to a reservoir. Yevjevich and Obeysekra (1985) studied the results of different drought characteristics and the correlations among them using generated data with various assumptions, such as the number of data generated, and using different lag 1 correlation coefficients. Guerrero-Salazer and Yevjevich (1975) analyzed the probability distribution function of duration and severity after identifying the wet and dry spells in a year.
- *Estimation of the pdf of drought characteristics* is a common method in hydrologic drought investigation. Drought studies were done by Sen (1980), who tried to use sequential periods (runs) as analytical tools to solve drought problems, such as the probability density function of drought characteristics. Yevjevich and Obeysekera (1985) and Yevjevich (1967) used generated data to determine the pdf of drought characteristics. Griffits (1990) determined the pdf of each drought characteristic and the multivariate distributions.

12.6.1 Run Theory

According to Yevjevich (1967), the basic method for local analysis of drought data is the application of run theory to determine the time series of hydrologic variables such as duration and severity of droughts. In this method, the following steps can be taken:

1. Annual streamflow data is separated into dry and wet years.
2. All sequential years in which the annual runoff is below long-term average are considered as dry events.
3. All sequential years in which the annual runoff is above long-term average are considered as wet events.

The above procedure is initialized on an annual time scale. However, if the time scale is less than one year and includes periodic parameters, a seasonal or monthly time scale can be used. Analysis of sequential periods helps the evaluation of the probability distribution of drought duration and severity. It should be noted that the time series usually are not long enough for this analysis; therefore, sufficient data must be generated to obtain a long-term time series. The bootstrap, tree ring, and autoregressive moving average (ARMA) methods have been used for data generation in various investigations.

12.6.2 Probability Distribution Function of Drought Characteristics

The final objective of analyzing a natural phenomenon is to quantify its behavior. This can usually be done by using the probability distribution function of the data. For this purpose, the pdf curves are obtained by frequency analysis of generated data after determining the duration and severity of droughts.

It is common to use the concept of return period for frequency analysis, but it is not suitable here as drought is not an annual event (there is not one specific observation per year). A long-term series (e.g., 500 years) is generated to estimate the pdf for various durations. Other questions that must be answered include:

- What is the largest drought duration and what is its severity?
- What is the most severe drought between the probable drought events?

Answering these basic questions requires a large number of statistical data series for which only long-term generated data are not sufficient; therefore, a planning horizon should be considered that is usually the same as the available historical data period or an existing planning horizon. The number of samples that must be generated is also determined. Because each sample represents only one value for each predefined objective (for example, each sample has only one drought with maximum intensity), the number of generated series should be large enough for frequency analysis.

12.6.3 Reservoir Operation and Demand Allocation

One of the most critical aspects of hydrologic drought is consideration of water demand. Reservoirs are usually used as the water supply system in large basins. The operating policies of existing reservoirs are an important tool for dealing with hydrologic droughts. Some researchers have studied the effects of reservoirs during drought (Chang et al., 1995). Optimal operating policy of a reservoir tends to allocate water to meet demands during droughts as much as possible.

12.7 AGRICULTURAL DROUGHT: DEFINITION AND BACKGROUND

In the study of *agricultural droughts*, many investigators believe that soil moisture deficit is the best and most practical drought indicator. Soil moisture shortage is the result of moisture loss during dry periods when evapotranspiration exceeds precipitation. Such losses depend on the type of plants. Once the soil moisture drops to a lower level than the plant roots, the actual evapotranspiration falls to a very low rate compared to the potential evapotranspiration. It is necessary to use water balance models to determine soil moisture shortage in different time periods, especially in regions with limited data availability.

Water balance models are based on theoretical methods to simulate runoff and soil moisture variations in a river basin. Water balance models were first developed by Thornthwaite (1948) and later modified by Thornthwaite and Mather (1955). These models include methods for estimating the water balance between the river basin inflow (from precipitation and snowmelt) and outflow (evapotranspiration and groundwater supply). Since development of the first water balance model by Thornthwaite and its subsequent modifications, these models have been used widely for irrigation management, prediction of runoff, and reservoir and groundwater supply (Haan, 1972; Howard and Lloyd, 1979; Mather, 1981; Van Bavel, 1953). Revelle and Waggoner (1983), Gleick (1987), and Schaake (1990) have also used water balance to estimate the effect of general climatic changes on hydrologic characteristics of specific regional models.

Water balance models are used to obtain a drought indicator. The Palmer drought severity index (PDSI) (Palmer, 1965) uses a water balance model between moisture supplied and demand for a two-layer soil. The PDSI is a tool for determining spatial and temporal variation in soil moisture stresses. Its calculation includes determination of Z index, the monthly soil moisture indicator that reflects the deviation of actual precipitation, and soil moisture supply from the expected values for normal conditions known to be climatically appropriate for existing conditions (*CAFEC*).

12.7.1 Thornthwaite Water Balance Model

The model initially developed by Thornthwaite (1948) is an old model and may be considered outdated, but it is still being used by many investigators (e.g., Alley, 1984), as is true for the PDSI. The Thornthwaite model is not data intensive and

works in regions with data deficiencies. Furthermore, it can be calibrated at different stages so that it is less susceptible to error accumulation. This model is based on factors such as existing soil moisture, soil moisture deficit, and total amount of rainfall and provides a general view of soil moisture variation resulting from variations in precipitation over the historical record. In other words, results of this model can be used for a general evaluation of current conditions caused by the impacts of regional climatic parameters on hydrologic and agricultural conditions. This water balance model is applicable to daily, monthly, or even annual time sequences.

The input data for the model are precipitation, monthly average temperature, and potential soil moisture. The model output consists of actual soil moisture, infiltration, excess rainfall, and actual and potential evapotranspiration. The Thornthwaite equation can be used to calculate the monthly potential evapotranspiration (*PE*) as follows:

$$PE = 16(10\theta / I)^a * f(\lambda) \tag{12.18}$$

$$a = 6.75 * 10^{-7} I^3 - 7.71 * 10^{-5} I^2 + 1.79 * 10^{-2} I + 0.49239 \tag{12.19}$$

$$I = \left(\theta / 5\right)^{1.514} \tag{12.20}$$

where:

PE is the potential evapotranspiration (mm).
θ is the average temperature for a specific time period (e.g., month) in ºC.
I is the annual temperature indicator, which is the summation of temperature indicator *i* over the 12 months of a year.
$f(\lambda)$ is the regional correction factor (a function of latitude and month).

The procedure for actual evapotranspiration and soil moisture calculation is as follows: When precipitation for a specific month is greater than the potential evapotranspiration, the actual evpotranspiration is equal to the potential evapotranspiration, and the remaining rainfall is compared with both the soil moisture capacity and the initial soil moisture of the previous month. The remaining rainfall supplies the soil moisture until it reaches its potential value. No runoff occurs until the soil approaches its moisture capacity. Obviously, if the summation of rainfall and soil moisture is less than the potential evapotranspiration, the actual evapotranspiration is less than the potential evapotranspiration.

12.7.2 Agricultural Drought Indicators

For trend studies, climatic factors such as temperature and precipitation are considered in addition to hydrological factors such as runoff (Thornthwaite, 1948). The following steps are taken to calculate the *moisture anomaly index* (*Z*), which is necessary for calculation of the PDSI.

12.7.2.1 Potential Evapotranspiration

As explained previously, the Thornthwaite equation is applied to calculate potential evapotranspiration. The advantage of this method is that temperature, the only required input parameter, is measured for most regions at acceptable time intervals. Other methods, such as the *T* model by Mather (1981) and *P* model by Maan (1972), require a number of climatic parameters that are seldom measured for one region at the same time intervals.

12.7.2.2 Soil Moisture Condition

The objective of this step is to identify the pattern of *soil moisture* variation as a function of temperature and precipitation. First, the precipitation at a specific month is compared with the potential evapotranspiration for that month. If the precipitation is sufficient, then the actual and potential evapotranspiration are assumed to be equal. After subtracting the potential evapotranspiration, the remaining amount will supply the soil moisture until it reaches its capacity (Φ). The rain that remains produces runoff and results in groundwater recharge. If the precipitation is less than potential evaporation, then the actual evapotranspiration could be provided from soil moisture from the previous month. The following equations express these situations:

$$\text{If } P_i \geq PE_i \text{ then: } \begin{cases} ET_i = PE_i \\ S_i = \text{Min}\left\{\left(P_i - PE_i\right) + S_{i-1}, \Phi\right\} \\ R_i = S_i - S_{i-1} \\ L_i = 0 \\ EP_i = (P_i - PE_i + S_{i-1}) - \Phi \end{cases} \tag{12.21}$$

$$\text{otherwise: } \begin{cases} ET_i = Min\left\{PE_i, (P_i + S_{i-1})\right\} \\ S_i = (P_i + S_{i-1}) - ET_i \\ R_i = 0 \\ L_i = S_{i-1} - S_i \\ EP_i = 0 \end{cases} \tag{12.22}$$

where:

P_i is the precipitation in month i.
PE_i is the potential evapotranspiration in month i.
ET_i is the actual evapotranspiration in month i.
S_i is the soil moisture content in month i.
R_i is the amount of water that the soil gained in month i.

L_i is the soil moisture loss in month i.
Φ is the soil moisture capacity.
EP_i is the excess rainfall in month i.

Then, the soil moisture deficit (D_i) is calculated as follows:

$$D_i = \Phi - S_i \tag{12.23}$$

In the first condition, the excess rainfall can be greater than 0, which means that the soil moisture deficit is equal to 0 and excess water is available for runoff.

Example 12.6

The monthly water balance parameters of a region are provided in Table 12.12. Assume that the soil moisture content does not change before and after rainfall. What is the evapotranspiration in that month (in mm)?

Solution: Direct runoff and groundwater recharge can be estimated as:

$$\text{Direct runoff} = 60 \times 0.3 = 18 \text{ mm}$$

$$\text{Groundwater recharge} = 100 \times 0.06 = 6 \text{ mm}$$

The actual evapotranspiration can then be estimated as the difference between rainfall depth and depth of direct runoff and groundwater recharge:

$$\text{Evapotranspiration} = 60 - 18 - 6 = 36 \text{ mm}$$

12.7.2.3 Runoff Calculation

Runoff is generated as a result of excess rainfall. The simplified water balance model is applied to calculate excess rainfall. Separation of runoff and groundwater recharge is difficult. Runoff can be calculated as follows:

$$RO_i = F^*EP_i \tag{12.24}$$

TABLE 12.12
Parameters of Water Balance Model for Example 12.6

Precipitation (mm)	Runoff Coefficient	Changes in Water Table (cm)	Storage Coefficient of Aquifer
60	0.3	+10	6%

where:

RO_i is the calculated runoff from excess rainfall in month i.
F is the lag coefficient for converting excess rainfall to runoff.
EP_i is excess rainfall in month i.

As it can be seen in this equation, a lag coefficient is necessary to calculate the runoff, which depends on the geophysical condition of a region. This factor is less than 1 and is calibrated by using long-term runoff data measured at gauging stations and long-term excess rainfall data calculated in the previous step. The method of least mean square error is used to calibrate the model and select the best lag factor for the region.

12.7.2.4 Potential Climatic Values

Besides the potential evapotranspiration, other potential values should be calculated as follows:

- *Potential recharge* (*PR*) is the moisture required by the soil to reach its moisture capacity. It is the difference between soil moisture capacity and the actual soil moisture which is the same as soil moisture deficit, D_i.
- *Potential moisture loss* (*PL*) is the amount of moisture lost from the soil to supply the evapotranspiration demand ($PE_i - S_{i-1}$). The moisture loss approaches its potential when there is no precipitation.

$$PL_i = \text{Min}\left\{PE_i, S_{i-1}\right\} \tag{12.25}$$

- *Potential runoff* (*PRO*) is the maximum runoff that can occur. It is a function of precipitation less the soil moisture gain. In the most critical conditions, it can be assumed that when the amount of runoff is high, the amount of potential recharge (*PR*) is low, and when runoff is low, the potential recharge is high. When *S* reaches its maximum value, the potential runoff is also maximized. The nature of potential runoff is much more complicated than other potential values. Different investigators have assumed values as high as 3 times the amount of precipitation and as ambiguous as the difference between soil moisture capacity and potential recharge. In the absence of evidence that could relate *PRO* to other hydrologic parameters, the value of *PRO* for each month is assumed to be the highest runoff experienced during the historical record for that month.

12.7.2.4.1 Coefficients of Water Balance Parameters

- *Evapotranspiration coefficient* (α): In dry climates, actual evapotranspiration is less than potential evapotranspiration. The amount of precipitation required for near-normal conditions is dependent on the average

climate and on the prevailing meteorological conditions during the period in question. α is used to estimate the expected evapotranspiration that is climatically appropriate for existing conditions (CAFEC) from the potential evapotranspiration. This terminology has been used frequently in drought studies. It simply implies the expected normal condition:

$$\alpha = \frac{\overline{ET_i}}{\overline{PE_i}} \tag{12.26}$$

where $\overline{ET_i}$ and $\overline{PE_i}$ are the averages of the historical *ET* and *PE*, respectively, in month *i*. The expected *ET* obtained by applying this coefficient could be compared with its actual amount in a specific month to obtain the deviation from normal conditions for that month.

- *Recharge coefficient* (β): In many regions, soil moisture recharge is a seasonal property of the soil. The proportion of average recharge for month *i* to average potential recharge is β:

$$\beta = \frac{\overline{R_i}}{\overline{PR_i}} \tag{12.27}$$

This coefficient could be used to compare the expected recharge in a specific climatic condition with its actual amount.

- *Runoff coefficient* (γ): Coefficient γ is calculated as the ratio of average runoff for month *i* to the potential runoff:

$$\gamma = \frac{\overline{RO_i}}{\overline{PRO_i}} \tag{12.28}$$

By applying this coefficient, the expected runoff in a specific climatic condition is calculated.

- *Moisture loss coefficient* (δ): Coefficient δ is the ratio of average moisture losses in month *i* to potential loss:

$$\delta = \frac{\overline{L_i}}{\overline{PL_i}} \tag{12.29}$$

The amount of soil moisture losses calculated by this coefficient could be subtracted from the sum of the other three factors to obtain the amount of precipitation for a region in normal condition.

12.7.2.4.2 Climatically Appropriate Precipitation for the Existing Condition (CAFEC) ($\hat{P}$)

Using the above equations, the amounts of evapotranspiration, recharge, runoff, moisture loss, and expected precipitation for each month under normal conditions can be calculated as follows:

$$\hat{ET}_i = \alpha.PE_i \tag{12.30}$$

$$\hat{R}_i = \beta.PR_i \tag{12.31}$$

$$\hat{RO}_i = \gamma.PRO_i \tag{12.32}$$

$$\hat{L}_i = \delta.PL_i \tag{12.33}$$

$$\hat{P}_i = \hat{ET}_i + \hat{R}_i + \hat{RO}_i - \hat{L}_i \tag{12.34}$$

Each component of CAFEC precipitation has an average equal to the average historical observations. The expected evapotranspiration, $\hat{ET}_i$, is equal to:

$$\hat{ET}_i = \alpha \cdot PE_i = \frac{\sum_{i=1}^{n}\left(ET_i\right)}{\sum_{i=1}^{n}\left(PE_i\right)} PE_i, \tag{12.35}$$

By taking the summation of both sides:

$$\sum_{i=1}^{n}\left(\hat{ET}_i\right) = \frac{\sum_{i=1}^{n}\left(ET_i\right)}{\sum_{i=1}^{n}\left(PE_i\right)} \sum_{i=1}^{n}\left(PE_i\right) \tag{12.36}$$

Therefore:

$$\sum_{i=1}^{n}\left(\hat{ET}_i\right) = \sum_{i=1}^{n}\left(ET_i\right) \tag{12.37}$$

and the average value of CAFEC evapotranspiration is equal to the average evapotranspiration from historical hydrologic accounting. The same reasoning holds for the averages of other components of CAFEC precipitation; therefore, the estimated value of CAFEC precipitation is unbiased. In spite of having the same averages, the CAFEC values and the actual/observed values for each period are seldom in complete agreement; however, in large period studies and observations they tend to converge.

12.7.2.4.3 Anomaly Moisture Index (Z)

After calculating CAFEC precipitation and comparing it with observed precipitation, the difference (d) can be calculated as:

$$d_i = P_i - \hat{P}_i \tag{12.38}$$

These departures from normal series show the deviation of weather from its normal condition. This new series provides a good description of the moisture or lack of moisture in the climate.

12.7.2.4.4 Climatic Character (K)

Average moisture supply is not always dependent on the precipitation in a period. Sometimes, when precipitation is insufficient, storage of moisture in the previous period is used. Using the average of moisture supplied, a climatic character (k_i) for each month i is calculated as follows (Palmer, 1965):

$$K_j = 17.67K'_j / \sum_{i=1}^{12} \bar{D}_i \times K'_j \qquad j = 1,\ldots 12 \tag{12.39}$$

where

$$K'_j = 1.5\log_{10}\left(\frac{T_j + 2.8}{\bar{D}_j}\right) + 0.5 \tag{12.40}$$

and

$$T_j = (\overline{PE}_j + \overline{R}_j + \overline{RO}_j) / (\overline{P}_j + \overline{L}_j) \tag{12.41}$$

T_j is a measure of the ratio of moisture demand to moisture supply for a month, and (K) is a regional climatic character. To estimate the Z index for a specific region, the coefficients of the above equations must be calculated based on the climatic character of the region. The monthly constants (K) are used as weighting factors of monthly deviations during dry spells. Then, a moisture anomaly index, Z, for month i is defined as follows:

$$Z_i = K \times d_i \qquad (i = 1,2,\cdots,N) \tag{12.42}$$

12.7.2.4.5 Converting the Z Index to a Drought Severity Index

To convert the Z index to a drought severity index, some regional studies are required. The Palmer studies were carried out in western Kansas, central Iowa, and the northwestern part of North Dakota. The PDSI introduced by Palmer for these regions is:

$$\begin{aligned} PDSI_i &= 0.897 \times PDSI_{i-1} + (1/3)Z_i \qquad i = 1, 2, \cdots, N \\ PDSI_1 &= (1/3)Z_1 \end{aligned} \tag{12.43}$$

The coefficients of the above equation depend upon the region and will change in various zones and regions with different climatic regimes. To obtain the above equation for a specific region, all sub-basins of the region must be investigated. The methodology of calculating the PDSI coefficients is described as follows. First, the Z index series, which will serve as the basis of the agricultural index, is calculated for all sub-basins. Based on the calculated Z index series for all sub-basins of the region, the following steps convert the Z index to a drought severity index:

1. For all negative values in Z index series, cumulative Zs are calculated.
2. Periods representing the maximum rates of the accumulated negative values are selected and plotted.
3. A straight line is drawn that indicates the approximate maximum rates observed during extremely dry periods of various lengths.
4. The ordinate from normal to extreme Z is divided into four equal lengths and numbered –1 to –4 to represent the severity of drought. The categories for drought severity and the related degrees of severity for these categories can be the same values used by Palmer (1965).
5. An equation between X (drought severity index), Z index, and t (duration) can then be obtained.

Karamouz et al. (2002) obtained the following equation for a specific region in the central part of Iran:

$$X_i = \sum_{t=1}^{i} Z_t \,/\, (5.14t + 10.88) \tag{12.44}$$

Extending the methodology proposed by Palmer (1965), Karamouz et al. (2002) derived the following equation for a specific region in the central part of Iran.

$$PDSI_i = 0.679 \times PDSI_{i-1} + (1/16)Z_i \qquad i = 1, 2, \cdots, N \tag{12.45}$$

TABLE 12.13
Amount of Abnormal Dryness Required To Maintain a Given Drought Severity

T	X_{i-1}	$\sum_{t=1}^{i-1} Z_i$	ΔX_i	X_i	$\sum_{t=1}^{i} Z_i$	Z_i
2	–1	–16.02	0	–1	–21.16	–5.14
10	–1	–57.14	0	–1	–62.28	–5.14
2	–3	–48.06	0	–3	–63.48	–15.42
10	–3	–171.42	0	–3	–186.84	–15.42

12.7.2.4.6 An Example of How to Derive the Drought Severity Index for a Region

Equation (12.44) is only an approximation of the drought severity index. The problem must be changed to be able to evaluate the severity of drought in monthly sequences. Such changes will provide the duration indirectly as a consequence of the accumulation of successive monthly contributions to drought severity. In order to evaluate the contribution of each month, we set $i = 1$ and $t = 1$ in Eq. (12.44) and we have:

$$X_1 = Z_1/16 \tag{12.46}$$

Because this is an initial month,

$$X_1 - X_0 = \Delta X_1 = Z_1/16 \tag{12.47}$$

A certain amount of dryness ($Z < 0$) is required to maintain the severity of the existing dry spell. Now we must determine this dryness for $\Delta X = 0$ (Table 12.13) (Palmer, 1965). The rate of increasing Z, in order to maintain a constant value of X ($\Delta X = 0$), depends on the value of X that is to be maintained (Palmer 1965). This reasoning adds additional term to Eq. (12.47), and the equation is changed to the form:

$$\Delta X_i = (Z_i/16) + c\, X_{i-1} \tag{12.48}$$

where:

$$\Delta X_i = X_i - X_{i-1}$$

Now the problem is to determine the c coefficient. Using Eq. (12.44), the value of z_i will maintain a given value of X from month to month. These values for two arbitrary values of $X_{i-1} = X_i$ and two arbitrary values are shown in Table 12.13. If we bring the values of Z_i, X_{i-1}, and ΔX from Table 12.13 into Eq. (12.48), we obtain:

$$\Delta X = 0 = (-5.14/16) - 1.0\ c \tag{12.49}$$

and

$$\Delta X = 0 = (-15.42/16) - 3.0\ c \tag{12.50}$$

Therefore, c is –0.321, and the final equation for the monthly contribution to drought severity is:

$$\Delta X_i = (Z_i/16) - 0.321\ X_{i-1}$$

In the other form, we have:

$$X_i = X_{i-1} + (Z_i/16) - 0.321^* X_{i-1} \tag{12.51}$$

Then, Eq. (12.45) is obtained.

12.7.2.4.6 Palmer Drought Severity Index

The PDSI is an index for evaluating the severity of a drought. After determining the Z index, the PDSI formulation is expressed as follows:

$$PDSI_i = 0.897\ PDSI_{i-1} + (1/3)\ Z_i \qquad i = 1,2,\ldots,\mathrm{N}$$
$$PDSI = (1/3)\ Z_1 \tag{12.52}$$

where N is the number of time periods. Dry and wet spells are categorized by the PDSI; see Table 12.14 for the drought classifications based on PDSI values.

TABLE 12.14
Drought Categories Based on PDSI

PDSI	Drought Category
≤–4	Most severe drought
–4 to –3	Severe drought
–3 to –2	Medium drought
–2 to –1	Nearly drought
–1 to 1	Normal
1 to 2	Nearly wet
2 to 3	Medium wet
3 to 4	Severe wet
≥4	Most severe wet

Source: Palmer, W. C., *Meteorological Drought*, U.S. Weather Bureau Research Paper No. 45, 1964.

To obtain drought and wet spells using the PDSI, one of the three different indices (X_i^k) must be used as follows:

$$X_i^k = 0.897 * X_{i-1}^k + (1/3) \times Z_i \quad k = 1,2,3 \quad i = 1,2,\ldots,N \tag{12.53}$$

where:

X_i^1 is the severity index for a wet spell in month *i*.
X_i^2 is the severity index for a dry spell in month *i*.
X_i^3 is the severity index for a spell that cannot be classified as dry or wet in month *i*.

To determine the drought severity index, the following steps should be taken: If X_i^2 is greater than or equal to –1, then month *i* is in a dry spell. If X_i^1 is greater than or equal to 1, then month *i* is in a wet spell. During a drought period, $X_i^3 = X_i^2$, and during a wet period $X_i^3 = X_i^1$. In this situation, when the indicators (X_i^1 or X_i^2) are between –0.5 and 0.5, then X_i^3 is equal to 0, which represents the termination time of a dry or wet spell, respectively. Often only one indicator is not equal to 0, and this indicator is the PDSI. The advantage of this method is that once a dry or wet period is observed by either X_i^2 or X_i^1, then the other resets to 0. Therefore, the value of the PDSI does not grow without any limits, which could be the case if only Eq. (12.52) is used (Alley, 1984).

The termination of an established drought is assumed to occur when $Z_i \geq Ze_i$, where Ze_i is the moisture required to reduce the severity of a drought to –0.5. Similarly, the termination of a wet spell is assumed to occur when $Z_i \leq Ze_i$. Ze_i can be derived by solving Eq. (12.52) for Z_i and substituting –0.5 and 0.5 for X_i, for dry and wet spells, respectively. The parameter introduced by Palmer (1965), *Pe*, is used to determine the correct value for the PDSI under various conditions. *Pe* is the probability that an established wet or drought spell has ended and is defined as:

$$Pe_i = \frac{100 \sum_{j=0}^{j^*} U_{(i-j)}}{Ze_i + \sum_{j=1}^{j^*} U_{(i-j)} - U_i} \tag{12.54}$$

where

$$U_i = Z_i + 0.15 \text{ for an established drought spell} \tag{12.55}$$

and

$$U_i = Z_i - 0.15 \text{ for an established wet spell} \tag{12.56}$$

Based on the study of Palmer (1965), in the case of an established drought a value of $Z = 0.15$ will maintain an index of –0.5 from month to month. Therefore, any value of Z greater than or equal to –0.15 will tend to end a drought. A similar assumption could be considered for the termination of a wet spell, and Palmer proposed the U_i as defined here.

The parameter j^* corresponds to the number of successive values of U_i computed prior to the current month. When the PDSI series shows a drought or wetness period, Pe must be at its extreme values (0 or 100) to change the current condition. When $0 < Pe < 50$, then the PDSI = X^3; when $50 < Pe < 100$, then the PDSI = X^2 or X^1, depending on which term is the opposite of the sign of $X3$. Readers are referred to Palmer (1965) and Alley (1984) for more details.

As mentioned in the methodology for calculating the PDSI, the coefficients of the Palmer drought severity index are dependent on the regional climate and may change with the climatic characteristics of various regions. So, it is necessary to derive new coefficients for the PDSI equation for different study areas.

Example 12.7

An example of the estimated PDSI for a specific region is provided in Figure 12.6, which shows the time series of the PDSI and standardized effective rainfall for the Zayandeh-rud River basin in the central part of Iran for a period of 34 years as monthly data from 1967 to 2002. Also, drought periods identified by the PDSI through the method described in Section 12.7 and effective rainfall calculations described in Section 12.4 are shown in the figure. The PDSI series shows seven drought periods, and the seventh period has not yet terminated. Effective rainfall series also show seven drought periods, and the seventh period has not yet terminated. Comparison of the two series reveals an overlap between the dry periods in both series. Similarities between the onset and termination times of both series show the relative consistency of both methods in identifying drought periods. Figure 12.6 shows the variations in standardized effective rainfall and PDSI time series and the drought spells identified by these methods.

12.8 GEOSTATISTIC TOOLS IN DROUGHT STUDIES

Geostatistics refers to the study of phenomena that fluctuate in space and/or in time. Geostatistics offers a collection of deterministic and statistical tools for modeling spatial and/or temporal variability. The basic concept of geostatistics is to determine any unknown value z as a random variable Z, with a probability distribution that models the uncertainty of z. The random variable is a variable that can take a variety of values according to a probability distribution. One of the most well-known geostatistic methods is kriging, which is widely used in the field of environment, mining, surveying, and water resources engineering studies. Geostatistics has been used in recent years for regional drought analysis. Because of their ability to characterize spatiotemporal events, these methods are efficient tools for drought studies, especially in the field of climatic drought assessment.

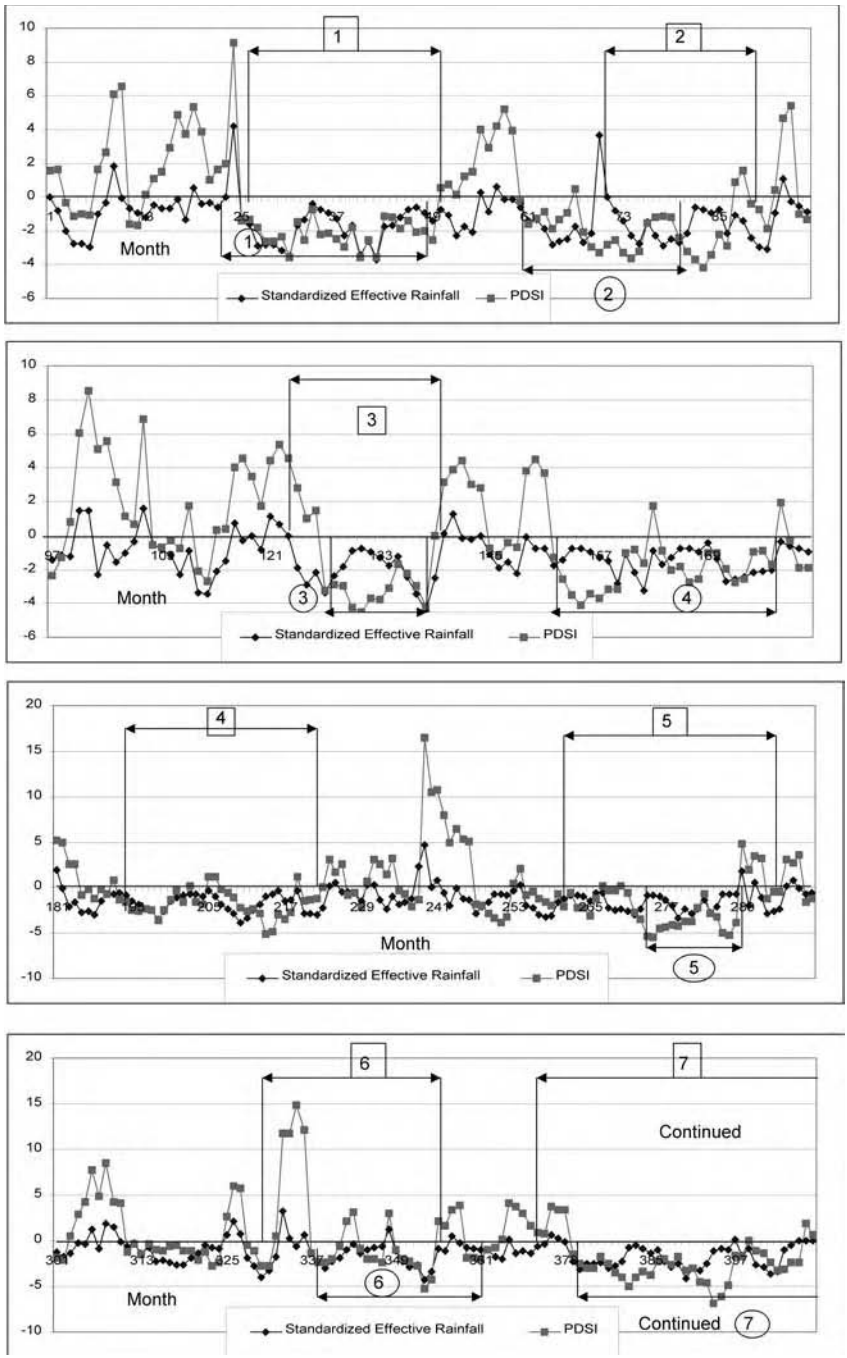


FIGURE 12.6 Dry and wet spells identification using PDSI (○) and effective rainfall method (□).

12.8.1 Kriging

Kriging is a collection of generalized linear regression techniques used to minimize estimation variance which is defined from a prior model for a covariance. It is important to have area-wide drought information before applying drought-control measures to an entire basin. Kriging can be used in its spatial form to determine the drought-affected areas in a region or for future installation of new gauging stations in ungauged areas. Kriging methods may be used to study the temporal variation of a process. In this case, a temporal domain, time variogram, and covariance function could be used to predict variables for unobserved data. Regional data resulting from dynamic processes in space and time use variogram and covariance functions that allow prediction of that process in different points in space and with different lead times.

Kriging in the simplest form applies first-order analysis based on the following intrinsic hypotheses (Bogardi et al., 1985):

- The expectation of $Z(x)$, a single realization of the parameter z, exists and is independent of point x:

$$E\left[Z(x)\right] = m \tag{12.57}$$

  Or if there is a known function $m(x)$ in which

$$E\left[Z(x)\right] = m(x) \tag{12.58}$$

  a variable such as $Y(x)$ could be used as:

$$Y(x) = Z(x) - m(x) \tag{12.59}$$

  And

$$E\left[Y(x)\right] = 0 \quad \text{for all } x \tag{12.60}$$

- For all vectors h, the increment $[Z(x + h) - Z(x)]$ has a finite variance independent of x:

$$Var\left[Z(x+h) - Z(h)\right] = 2\gamma(h) \tag{12.61}$$

  The γ function is the *variogram*. Variograms are obtained from the following three parameters: (1) *nugget effect*, which can partly be explained by the measurement error (C_0); (2) *sill*, which is the total variance of the data beyond a certain distance or range (C); and (3) *range* or *zone of influence* (R) (see Figure 12.7).

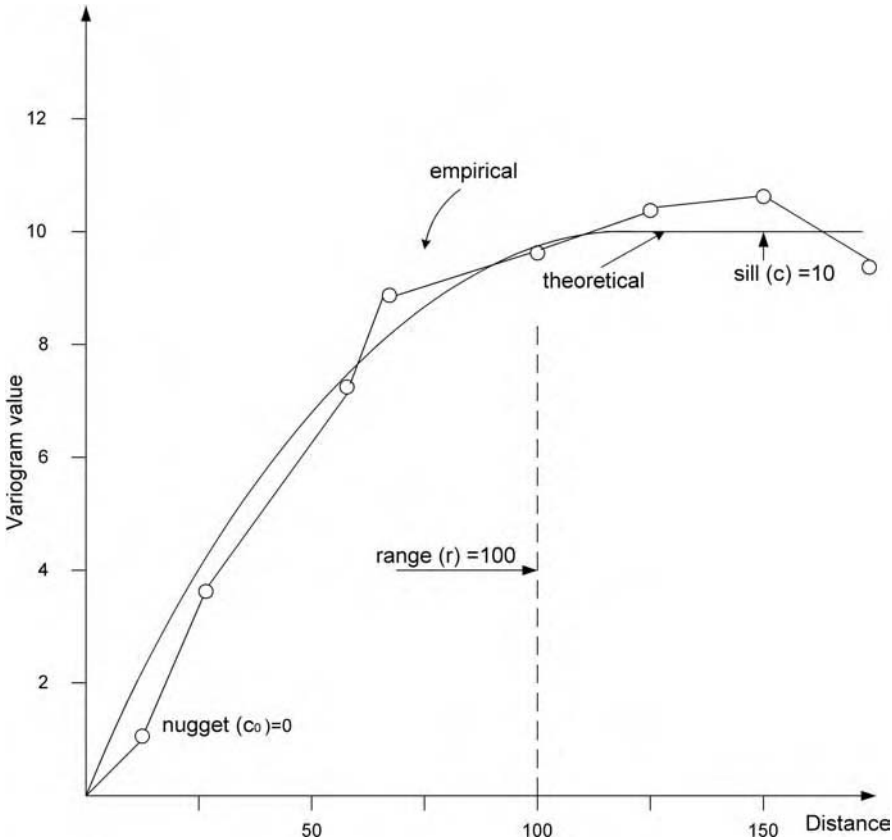


FIGURE 12.7 An example of empirical and theoretical variogram.

The first step in kriging analysis is to estimate a semivariogram. In all applications, the purpose is to analyze the correlation between neighboring measurements of a property and thereby derive the spatial variability of such a property. The functional description of this spatial structure, if it exists, is known as the semivariogram in geostatistics. The computational form of a semivariogram is:

$$\gamma(h) = \frac{1}{2n} \sum \left[\left(z(x_i) - z(x_i + h) \right)^2 \right] \tag{12.62}$$

where $\gamma(h)$ is the semivariogram value for distance h, $z(x_i)$ is the measurement of location x_i, and $z(x_i + h)$ is the measurement at location $x_i + h$, where n is the number of measurement pairs separated by vectorial distance h.

The empirical variogram could be replaced by a theoretical variogram model to be used in a kriging problem. A theoretical semivariogram model consists of an isotropic nugget effect and any positive linear combination of the following standard semivariogram models:

- *Spherical model* (*Sph*(.)):

$$\gamma(h) = c.Sph\left(\frac{h}{a}\right) = \begin{cases} c.\left[1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right], if(h \le a) \\ c, if(h \le a) \end{cases} \tag{12.63}$$

where a is an actual range and c is the positive variance contribution or sill.

- *Exponential model* (*Exp*(.)):

$$\gamma(h) = c.Exp\left(\frac{h}{a}\right) = c.\left[1 - \exp\left(-\frac{3h}{a}\right)\right] \tag{12.64}$$

- *Gaussian model:*

$$\gamma(h) = c.\left[1 - \exp\left(-\frac{(3h)^2}{a^2}\right)\right] \tag{12.65}$$

- *Power model:*

$$\gamma(h) = c.h^{\omega} \tag{12.66}$$

where ω is a constant between 0 and 2.

- *Hole effect model:*

$$\gamma(h) = c.\left[1.0 - \cos\left(\frac{h}{a}.\pi\right)\right] \tag{12.67}$$

Figure 12.7 shows a schematic of empirical and theoretical variograms. Consider the estimate of an unknown (unsampled) value $z(u)$ from neighboring data values $z(u_a)$ ($a = 1, \ldots, n$). The random function model $Z(u)$ is stationary with mean m and covariance $C(h)$. In its simplest form, also known as simple kriging, the algorithm considers the following linear estimator:

$$Z^*_{SK}(u) = \sum\nolimits_{\alpha=1}^{n} \lambda_\alpha(u) Z(u_\alpha) + \left((1 - \sum \lambda_\alpha(u)) \right) m \tag{12.68}$$

where Z^*_{SK} is a simple kriging estimator, and $\lambda_\alpha(u)$ are the weights to minimize the error variance, also referred to as the *estimation variance*. Minimization of the error variance in a set of normal equations results:

$$\sum_{\beta=1}^{n} \lambda_\beta(u) C(u_\beta - u_\alpha) = C(u - u_\alpha) \quad \forall \alpha = 1, ..., n \tag{12.69}$$

where C is the stationary covariance and $C(0)$ is the stationary variance of $Z(u)$.

The corresponding minimum estimated variance, or *kriging variance*, is:

$$\sigma^2_{SK}(u) = C(0) - \sum_{\alpha=1}^{n} \lambda_\alpha(u) C(u - u_\alpha) \geq 0 \tag{12.70}$$

Ordinary kriging is the most commonly used method of kriging, whereby the sum of weights

$$\sum_{\alpha=1}^{n} \lambda_\alpha(u)$$

is limited to 1. This allows building an estimator $Z^*_{OK}(u)$ that does not require prior knowledge of the stationary mean m and remains unbiased in the sense that:

$$E\left\{Z^*_{OK}(u)\right\} = E\left\{Z(u)\right\} \tag{12.71}$$

The relationship between weights λ and the semivariograms introduced earlier can be represented in a matrix form as follows:

$$\begin{pmatrix} 0 & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & 0 & \gamma_{23} & \cdots & \gamma_{2n} & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \cdots & 0 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_0^1 \\ \lambda_0^2 \\ \vdots \\ \gamma_0^n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{pmatrix} \tag{12.72}$$

where λ_0^i represents the weight of station i, and 0 represents the origin station.

12.8.2 Spatial Kriging in Drought Studies

An important factor that should be considered in drought regionalization is the drought-affected area of the entire region. Kriging can be used in its spatial form to determine the drought-affected areas in a region. Also, spatial kriging errors could be used as guidance for future installation of new stations in an ungauged area. To determine the drought-affected areas in a region, these steps should be followed:

- A spatial network with specific grids is considered over the entire region.
- The observed data of all rain gauges are considered as known measured data at their nearest nodes of the network.

- The spatial variogram of precipitation is calculated and is fitted by a theoretical variogram.
- The unknown precipitation at each node of the network is obtained using the characteristics of the above variogram by a kriging method.
- As defined by regional criteria, drought-affected areas are estimated by the summation of all drought-affected subareas that have experienced precipitation less than the defined criteria.

An example of spatial kriging in regional drought assessment has been presented in Section 12.4 of this chapter.

12.8.3 Spatiotemporal Kriging

The behavior of environmental variables such as seasonal rainfall patterns can usually be explained by time-dependent patterns and require different water supply policies. These policies will be more important during drought events. A spatiotemporal study of precipitation over a region makes it possible to determine space and time variations of precipitation in a region. Also, simulation methods in geostatistics can be used to simulate realizations of precipitation time series at each location of the region. The realizations are useful for calculating the probable characteristics of a climatic drought (severity and/or duration) at each zone of a region at a given significant level.

The spatiotemporal time series could be studied through the assessment of a trend model and the residuals of precipitation time series at each station of a region. A time series at each station is divided by deterministic and random components as follows:

$$Z(\mathbf{u},t) = M(\mathbf{u},t) + R(\mathbf{u},t), \qquad \forall \mathbf{u} \in D, \qquad \forall t \in T \tag{12.73}$$

$Z(\mathbf{u}, t)$ is a set of rainfall time series during period T at each location $\mathbf{u}$, which is a two-dimensional vector; D represents the two-dimensional domain of the study area; $M(\mathbf{u}, t)$ is a deterministic spatiotemporal component that models the average variation of rainfall; and $R(\mathbf{u}, t)$ is a random variable component, independent from $M(\mathbf{u}, t)$, that models the fluctuations with higher frequency around the average spatiotemporal variations. Deterministic models of rainfall variations, $M(\mathbf{u}, t)$, are developed independently at each station $\mathbf{u}$. The stochastic model of spaciotemporal variation is then obtained by regionalizing the parameters of the deterministic models. The first goal at each station is to determine the periodic behavior of the rainfall data, such as inter-annual variations, which can be obtained through time series models such as spectral analysis.

Deterministic variations at each station $\mathbf{u}_\alpha$ can be modeled as the sum of $k + 1$ unknown temporal basic functions $f_k(t)$ as follows:

$$M(\mathbf{u}_\alpha, t_i) = \sum_{k=0}^{K} b_k(\mathbf{u}_\alpha) f_k(t_i), \qquad i = 1, \cdots, T_\alpha \tag{12.74}$$

where $b_k(\mathbf{u}_\alpha)$ is the coefficient of the kth function. These functions usually consist of a linear function to show the long-term trend and the seasonal or annual periodic components. The periodic nature of the basic functions is determined as a component of $M(\boldsymbol{u},t)$ by a time series analysis method such as Fourier analysis. The coefficients of the model are regionalized by the kriging method. This makes it possible to predict the rainfall trend at each location of the region. Also simulation of the regionalized trend is possible using geostatistical simulation methods (Deutsch and Journel, 1999).

The residual values $R(\boldsymbol{u},t)$ are obtained by subtracting the average variations $M(\boldsymbol{u},t)$ from rainfall values, $Z(\boldsymbol{u},t)$ at each station. The model of the spatio-temporal residual, $R(\boldsymbol{u},t)$, is developed using spatially correlated residual time series.

Regionalizing and simulating data using various kriging and simulation methods could be applied to both trends and residuals of time-series modeling. Most geostatistics methods require time series to be in a Gaussian distribution environment; therefore, transformation of the data to normal values will be necessary in most cases.

The most straightforward algorithm for generating the realizations of a multivariate Gaussian field is provided by the sequential Gaussian simulation (Deutsch and Journel, 1998). Simulation through LU (lower and upper triangular matrices) decomposition of the covariance matrix is another method in geostatistical simulation.

The final real precipitation time series of each realization is obtained by combining the results of the simulated trend and residual models for all points (nodes) of the region.

Example 12.8

Consider a region with 36 rain gauges (Figure 12.8) which have recorded 50 years of precipitation data. The rain gauges are shown as points in a 30 by 30-unit network. Set up an algorithm for monthly space–time regionalization of the precipitation time series:

Solution: The algorithm is expressed by following steps:

1. At first, the real values of precipitation are transformed to normal values. The trend model for each station is obtained by fitting time series models such as spectral model. If we use a model with n parameter for all stations, $36 \times n$ parameters will be produced at this step.
2. Using the coefficients of trend model and sequential time steps, the precipitation trend series are obtained at each station, which are equal to 36 vectors of 600 (50×12) monthly precipitation trend values.
3. The residual time series at each station is obtained by subtracting the trend model from the real values of precipitation. This will produce 36 vectors of 600 precipitation residual values.
4. The parameters of trend model are regionalized using kriging methods. The results are presented as 900 (30×30) n values.
5. If m realizations are needed, the simulation of regionalized parameters should be done m times using simulation methods.

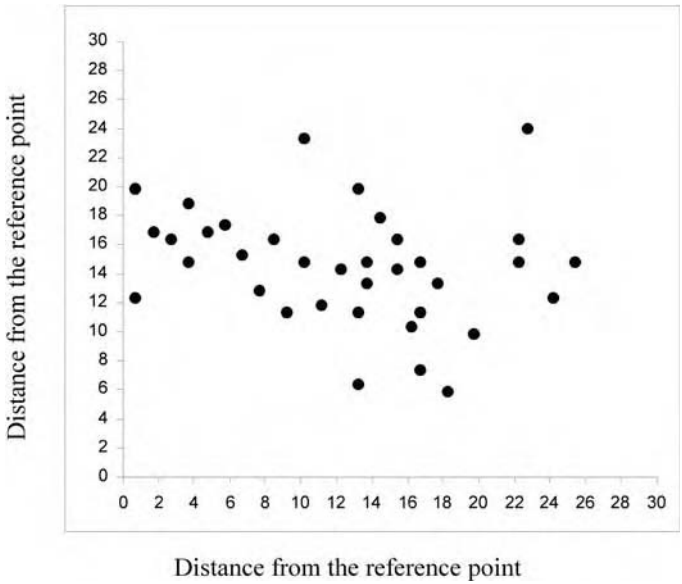


FIGURE 12.8 The location of rain gages in the region.

6. Using the coefficients of trend model and sequential time steps, the precipitation trend series are obtained at each station, which are equal to $900 \times m$ vectors of 600 values.
7. The precipitation residual time series at each station are regionalized using geostatistical methods. This will provide 900 residual time series of 600 values.
8. To obtain m realizations, m simulation of residuals should be done. Then the output will contain $900 \times m$ vectors of 600 values.
9. Compose the trend values and residuals at 900 nodes of the network to provide $900 \times m$ vectors of 600 values of the normalized precipitation time series.
10. Use inverse transformations to convert normalize data to real values.
11. Finally the result consists of m realizations of 50-year (600-month) time series at each 900 nodes of the network.

12.9 PROBLEMS

12.1 Use the 3-year moving averages to compare the two series of annual average precipitation for two different regions presented in the following table:

No. of Years	Precipitation of Region A (mm)	Precipitation of Region B (mm)
1	50	71
2	3	40
3	8	118
4	59	99
5	58	48
6	141	134
7	30	107
8	22	92
9	41	28
10	47	32
11	103	45
12	23	53
13	64	66
14	46	142
15	129	153
16	82	59
17	128	46
18	89	121
19	17	11
20	15	191
21	28	101
Current year	15	45

(a) Which of these regions is facing the most severe drought in the current year?
(b) If the long-term average precipitation in regions A and B is estimated to be 350 and 250 mm, respectively, discuss the drought periods and their durations.
(c) Use Eq. (12.13) to estimate the drought severity for each of the drought periods.

12.2 The monthly rainfall data for 22 years is presented in the following table. Assume C (the constant in Eq. (12.6)) is 0.1 and calculate the following items for the region:
(a) Effective rainfall
(b) Rainfall deficit
(c) Excess rainfall deficit.

12.3 Determine the onset, duration, and severity of drought for the first year of data presented in Problem 12.2.

12.4 The historical annual precipitation in a basin and the annual rainfall in drought events with different return periods are shown in the following

No. of Years	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.
1	0.2	49.6	71.0	40.9	77.1	109.2	153.3	40.9	9.8	0.7	1.6	0.4
2	0.4	2.9	40.3	106.8	73.7	90.8	83.5	5.0	0.0	15.9	0.8	0.0
3	10.4	8.3	117.7	88.4	67.2	88.7	59.8	12.8	0.0	1.4	0.0	0.0
4	0.0	59.2	98.6	205.1	179.7	76.5	37.4	73.1	13.6	0.0	0.6	0.0
5	8.8	57.9	47.5	124.3	28.1	77.7	124.5	27.8	1.4	1.4	6.7	0.2
6	38.0	140.7	133.9	54.5	55.6	128.8	82.4	45.5	14.2	0.3	2.0	0.0
7	2.5	29.5	106.6	48.1	43.3	31.7	15.4	31.5	0.3	0.3	0.0	0.0
8	4.7	22.0	91.8	61.2	164.8	159.9	83.7	97.4	8.7	0.3	1.3	1.4
9	1.9	40.5	27.5	122.9	74.9	73.2	93.9	23.1	0.0	0.4	0.0	0.0
10	15.3	46.8	31.7	62.6	80.1	121.3	125.4	27.5	0.1	0.4	0.0	0.0
11	50.2	102.7	44.6	93.1	58.3	107.0	65.1	28.0	9.6	0.0	0.0	0.0
12	0.0	22.6	53.2	63.0	40.8	117.3	39.2	55.3	2.8	0.0	0.4	0.1
13	17.4	64.4	66.2	59.5	64.3	55.5	92.3	49.9	0.2	0.0	0.0	0.0
14	0.5	46.4	142.0	24.2	67.2	90.2	41.8	29.5	0.9	0.0	0.1	1.1
15	0.5	128.8	153.2	73.5	69.6	175.7	73.3	95.6	2.9	0.0	0.6	0.5
16	33.4	82.2	58.5	151.7	99.0	145.9	57.1	38.1	0.2	10.6	14.9	0.0
17	16.3	127.5	45.7	67.7	56.6	82.2	24.9	34.5	0.3	0.9	0.2	0.2
18	0.0	88.7	120.8	49.2	74.0	115.2	41.4	51.6	0.2	0.0	0.0	0.0
19	2.1	17.2	10.7	77.8	65.4	116.4	47.0	4.8	0.0	0.9	0.0	0.0
20	47.7	14.8	190.6	82.5	48.0	110.6	97.8	1.9	0.4	1.0	0.0	0.3
21	2.3	28.2	101.3	139.0	131.2	252.9	59.4	96.9	5.5	0.7	0.0	0.5
22	0.7	77.2	45.3	74.7	78.2	102.9	59.1	135.8	0.0	0.0	0.3	1.1

tables. If a 25-year climatic drought occurs in the region, determine which sub-basins face droughts with 10-year or more return periods.

Regional Precipitation Data for Drought Events with Different Return Periods (Problem 12.4)

Sub-Basin	Area (%)	Drought Precipitation (mm)				
		2-Year	5-Year	10-Year	25-Year	50-Year
A	20	380	280	180	100	70
B	30	600	540	510	490	350
C	15	550	520	480	320	250
D	15	280	250	200	180	70
E	20	400	350	310	280	190

Annual Average Precipitation in a Basin (Problem 12.4)

No. of Years	Precipitation (mm)	Sorted Precipitation (mm)	Probability	Return Period (yr)
1	618.23	101.47	0.05	19.00
2	563.54	284.2	0.11	9.50
3	546.6	307.44	0.16	6.33
4	501.51	345.81	0.21	4.75
5	480.41	350.1	0.26	3.80
6	451.88	368.51	0.32	3.17
7	418.34	369.99	0.37	2.71
8	403.34	403.34	0.42	2.38
9	369.99	418.34	0.47	2.11
10	368.51	451.88	0.53	1.90
11	350.1	480.41	0.58	1.73
12	345.81	501.51	0.63	1.58
13	307.44	546.6	0.68	1.46
14	284.2	563.54	0.74	1.36
15	776.02	617.04	0.79	1.27
16	1001.96	618.23	0.84	1.19
17	101.47	776.02	0.89	1.12
18	617.04	1001.96	0.95	1.06

12.5 The data presented in the following table have been recorded in the region presented in Problem 12.4 for one specific year. What is the appropriate return period of regional drought?

	Station				
	A	B	C	D	E
Annual precipitation (mm)	310	490	350	275	390

12.6 Consider a basin with 12 rain gauges, the locations of which are shown in the following figure. The recorded rainfall at each station for a year are presented in the following table. Assume a threshold of 25-mm precipitation for the return period of a 10-year drought in the region. Specify the 10-year drought-affected areas in the region.

	Rain Gauges											
	1	2	3	4	5	6	7	8	9	10	11	12
Precipitation	320	315	205	105	120	240	230	231	242	250	317	318

References

Alley, W. M., The Palmer drought severity index: Limitation and assumptions" *J. Climate Applied Meteorol*., 23, 1100–1109, 1984.

Aswathanarayana. U., *Water Resources Management and the Environment*, Balkema Publishers, New York, 2001.

Beric, M., Zelenhasic, E., and Srdjevic, B., Extreme dry weather intervals of the growing season in Backa, Yugoslavia, *Water Resources Manage.*, 4, 37–73, 1990.

Ben–Zvi, A., Indices of hydrological drought in Israel, *J. Hydrol.*, 22, 179–191, 1987.

Bogardi, I., Bardossy, A., and Duckstein, L., Multicriterion network design using geostatistics, *Water Resources Res.*, 21(2), 199–208, 1985.

Chang, G., Effects of droughts on the streamflow characteristics, *ASCE J. Irrigation Drainage Eng.*, 116(3), 332–341, 1990.

Chang, T. J., Xenia, A. and Choo B. T., Use of Flood-Control Reservoirs for Drought Management" *ASCE J. Irrigation Drainage Eng*., 121(1), 34–42, 1995.

Deutsch C. V., Journel, A. G., *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 1998.

Dracup, J. A., Lee, K. S., and Paulson, E. G., On the definition of droughts, *Water Resources Res.*, 16(2), 297–302, 1980.

Gleick, P. H., Regional hydrologic consequences of increases in atmospheric CO_2 and other trace gasses, *Climatic Change*, 10(2), 137–161, 1987.

Griffits, G. A., Rainfall deficits: distribution of monthly runs, *J. Hydrol.*, 115, 219–229, 1990.

Guerrero–Salazer, P. L. A. and Yevjevich, V., *Analysis of Drought Characteristics by the Theory of Run*, Hydrology Paper 80, Colorado State University, Fort Collins, CO, 1975.

Haan, C. T., A water yield model for small watersheds, *Water Resources Res.*, 8(1), 58–69, 1972.

Herbst, P. H., Bredenkamp, D. B., and Barker, H. M. G., A technique for the evaluation of drought from rainfall data. *J. Hydrol.,* 4, 264–272, 1966.

Howard, K. W. F. and Lloyd, J. W., The sensitivity of parameters in the Penman evaporation equations and direct recharge balance, *J. Hydrol.*, 41, 329–344, 1979.

Karamouz, M., *Drought Management in Zayandeh-rud River Basin*, final report to Regional Water Authority of Isfahan, Iran, 2003.

Karamouz, M., Torabi, S., Araghi-Nejad, Sh., and D. Asgharzadeh, Analysis of droughts: application to isfahan region in Iran, *Proc. 2002 EWRI Conference on Water Resources Planning and Management*, Roanoke, VA, 2002.

Kottegoda, N. T. and Rosso, R., *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, McGraw-Hill, New York, 1998.

Lettenmiar, D. P., McCabe. G., and Stakhiv. E. Z, Global climate change: Effects on hydrologic cycle, in *Water Resources Handbook*, L. W. Mays, Ed., McGraw-Hill, New York, 1996.

Maan, C. T., A water yield model for small watersheds, *Water Resources Res.*, 8(2), 58–69, 1972.

Mather, J. R., Using computed stream flow in watershed analysis, *Water Resources Bull.*, 17(3), 474–482, 1981.

Mendenhall, W., Wackerly D. D., and Scheaffer, R. L., *Mathematical Statistics with Applications*, PWS-KENT, Boston Publishing Company, Boston, MA, 1990.

Palmer, W. C., *Meteorological Drought*, U.S. Weather Bureau Research Paper No. 45, 1965.

Revelle, R. R. and Waggoner, P. E., Effects of a carbon dioxide-induced climatic change on water supplies in the western United States, *Changing Climate*, National Academy Press, Washington D.C., 1983.

Rossi, G., Characteristics of drought over a region and shortage control strategies, *Proc. of the 18th IAHR Congress*, Cagliari, Vol. 2, 1979.

Rossi, G., Benedini, M., Tsakiris, G., and Giakoumakis, S., On regional drought estimation and analysis, *Water Resources Manage.*, 6, 249–277, 1992.

Rouhani, S. and Cargile, K. A., A geostatistical tool for drought management, *J. Hydrol.*, 108, 255–266, 1989.

Santos, M. A., Regional droughts: a stochastic characterizations, *J. Hydrol.*, 66, 183–211, 1983.

Schaake, J. C., From climate to flow, in *Climate Change and U.S. Water Resources,* Waggoner, P. E., Ed., Wiley Series in Climate and the Biosphere: Report of the American Association for the Advancement of Science Panel on Climatic Variability, Climate Change and the Planning and Management of U.S. Water Resources, John Wiley & Sons, New York, 1990, pp. 177–207.

Sen, Z., Regional drought and flood frequency analysis: theoretical consideration, *J. Hydrol.*, 46, 1980.

Thornthwaite, C. W., An approach toward a rational classification of climate, *Geogr. Rev.*, 38(1), 55–94, 1948.

Thornthwaite, C. W. and Mather, J. R., The water balance*, Publ. in Climatol. Lab. Climatol.*, 8, 1–86, 1955.

Van Bavel, C. H. M., A drought criterion and its application in evaluating drought indices and hazard, *Agron. J.*, 45(4), 167–172, 1953.

Yevjevich, V., *An Objective Approach to Definitions and Investigations of Continental Hydrologic Drought*, Hydrology Paper No. 23, Colorado State University, Fort Collins, CO, 1967, 18 pp.

Yevjevich, V. and Obeysekera, J. T. B., Correlation between sample first autocorrelation coefficient and extreme hydrologic runs, *J. Hydrol.*, 79, 171–186, 1985.

Zucchini, W. and Adamson, P. T., On the application of the bootstrap to assess the risk of deficient annual inflows to a reservoir, *Water Resources Manage.*, 2, 245–254, 1988.

Index

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Y

Z