ASCE Manuals and Reports on Engineering Practice No. 28

Hydrology Handbook

Second Edition

Prepared by the Task Committee on Hydrology Handbook
of Management Group D of the American Society of Civil Engineers

This is a completely re-written edition
of the Manual published in 1949.

Published by

ASCE

345 East 47th Street
New York, NY 10017-2398

Abstract:

This new edition of the *Hydrology Handbook* (Manual No. 28) incorporates the many changes and advances that have occurred in the areas of planning, development, and management of water resources since the publication of the original manual in 1949. The first six chapters, Chapters 2 through 7, relate to the natural phenomena in the hydrologic cycle, while the next three chapters describe the predictions and effects of the phenomena previously described. The final chapter reviews the applications of hydrology starting with study formulation, then reviews data management, then discusses calibration and verification of hydrologic models, and concludes with accessing accuracy and reliability of results. With this new edition, academic and practicing hydrologists have a thorough and up-to-date guide to the field of hydrologic engineering.

Library of Congress Cataloging-in-Publication Data

Hydrology handbook / prepared by the Task Committee on Hydrology Handbook of Management Group D of the American Society of Civil Engineers. -- 2nd ed.
p. cm. -- (ASCE manuals and reports on engineering practice ; no. 28)
Includes bibliographical references and index.
ISBN 0-7844-0138-1
1. Hydrology. I. American Society of Civil Engineers. Task Committee on Hydrology Handbook. II. Series.
GB661.2.H93 1996 96-41049
551.48--dc20 CIP

Library of Congress Catalog Card No: 96-41049
ISBN 0-7844-0138-1
Manufactured in the United States of America.

CONTENTS

CONTENTS

Chapter 1

INTRODUCTION TO THE NEW HANDBOOK OF HYDROLOGY

I. HISTORICAL SUMMARY

In 1949, the American Society of Civil Engineers (ASCE) published a hydrology handbook titled *Manual of Engineering Practice No. 28*, which was prepared by a Committee of the Hydraulics Division. It was reprinted in 1952 and 1957. When consideration was given to reprinting *Manual 28*, the matter was referred to a Task Committee of the Hydraulics Division in 1982. After considerable discussion and review of other publications, it was recommended that *Manual 28* not be reprinted but that a new handbook on hydrology be prepared. This new publication would recognize the advances that had been made in the practice of engineering hydrology in the years since *Manual 28* was published.

With the encouragement of the Executive Committees of the Hydraulics Division and the Irrigation and Drainage Division, the Task Committee prepared a table of contents for a new handbook of hydrology and presented it along with the recommendation that the new handbook be a cooperative effort of the two Divisions. In 1986, this recommendation was approved by the Executive Committees and was transmitted to Management Group D, the Group that oversees the activities of the five ASCE technical divisions involved in water engineering. Under the leadership of Conrad G. Keyes, Jr., Management Group D accepted the report and in 1988 established a Task Committee comprising four members, two from the Hydraulics Division and two from the Irrigation and Drainage Division, to prepare a new handbook of hydrology.

II. PURPOSE OF THE NEW HANDBOOK

The 1949 hydrology handbook, *Manual 28*, has been widely distributed and used, both nationally and internationally, by educational institutions, consulting engineers, and others involved in planning, development, and management of water resources. The manual has served essentially as a baseline document; however, since its publication many changes and advances have occurred, not only in a fundamental understanding of hydrologic systems, but also in data acquisition and computation techniques and procedures. These changes and advances have been incorporated into the new *ASCE Handbook of Hydrology* so that it maintains its relevancy to academic and practicing hydrologists throughout the world. Providing a thorough and up-to-date guide for hydrologists is the purpose of this new handbook.

III. SCOPE OF THE NEW HANDBOOK

The first six chapters, Chapters 2 through 7, relate to the natural phenomena in the hydrologic cycle. Chapter 2 describes the formation and types of precipitation, variations in and measurement of precipi-

tation, processing and interpreting precipitation records, frequency analyses of precipitation data, weather modification, and synthetic weather generation. Chapter 3 covers principles of infiltration, factors affecting infiltration and rainfall excess, infiltration and excess models, and measurement of infiltration. Chapter 4 presents the physics and theory of evaporation, interaction of surface and meteorological factors, evaporation from water surfaces, evapotranspiration from land surfaces, and regional evapotranspiration. Chapter 5 discusses ground water source and occurrence, subsurface medium, ground water movement, basin yield concepts and evaluations, recharge, ground water quality, models, and management. Chapter 6 describes the runoff process and its variability, measurement of stream flow, hydrographs, overland flow, stream flow routing, reservoir storage-yield analysis, and runoff quality. Chapter 7 covers the physical processes of snow and snowmelt; data requirements, collection, and sources; snow accumulation and distribution; snowmelt analysis and simulation; water supply forecasting; computer programs; and sample applications.

The next three chapters, Chapters 8 through 10, describe the predictions and effects of the phenomena described in the first six chapters. Chapter 8 presents flood characteristics and analysis, statistical analysis for estimating floods, estimating floods from rainfall, probable maximum flood, flood hazard, flood warning, and microcomputer software for flood analyses. Chapter 9 reviews the hydrologic impacts of urbanization, precipitation in the urban watershed, hydrologic losses in developing watersheds, urban runoff estimating methods, typical urban drainage design calculations, and computer model applications. Chapter 10 discusses wave theory, wind waves, ship-generated waves, wave-structure interaction, waves and currents, tides and tidal datums, storm surges, basin oscillations and tsunamis, water surface probability analysis, and selection of design water waves and levels.

The last chapter, Chapter 11, reviews the applications of hydrology starting with study formulation, then reviews data management, then discusses calibration and verification of hydrologic models, and ends with assessing accuracy and reliability of study results.

IV. THE HYDROLOGIC CYCLE

The constant movement of water and its change in physical state on this planet is called the water cycle, also known as natures waterwheel, or the hydrologic cycle. This cycle is depicted in Fig. 1.1. The word *cycle* implies that water derives from one source and eventually returns to that source. Water originates from the oceans and returns to the oceans. On its way, water may change its state from vapor (gas), to liquid (water), to solid (ice and snow) in any order.

A description of the hydrologic cycle can begin at any point and return to that same point. Water in the ocean evaporates and becomes atmospheric water vapor. Some of this moisture in the atmosphere falls as precipitation, which sometimes evaporates before it can reach the land surface. Of the water that reaches the land surface by precipitation, some may evaporate where it falls, some may infiltrate the soil, and some may run off overland to evaporate or infiltrate elsewhere or to enter streams. The water that infiltrates the ground may evaporate, be absorbed by plant roots and then transpired by the plants, or percolate downward to ground water reservoirs. Water that enters ground water reservoirs may either move laterally until it is close enough to the surface to be subject to evaporation or transpiration; reach the land surface and form springs, seeps or lakes; or flow directly into streams or into the ocean. Stream water can accumulate in lakes and surface reservoirs, evaporate or be transpired by riparian vegetation, seep downward into ground water reservoirs, or flow back into the ocean, where the cycle begins again.

Each phase of the hydrologic cycle provides opportunities for temporary accumulation and storage of water, such as snow and ice on the land surface; moisture in the soil and ground water reservoirs; water in ponds, lakes, and surface reservoirs, and vapor in the atmosphere. Without replenishment from precipitation, the water stored on all of the continents would gradually be dissipated by evapotranspiration processes or by movement toward the oceans.

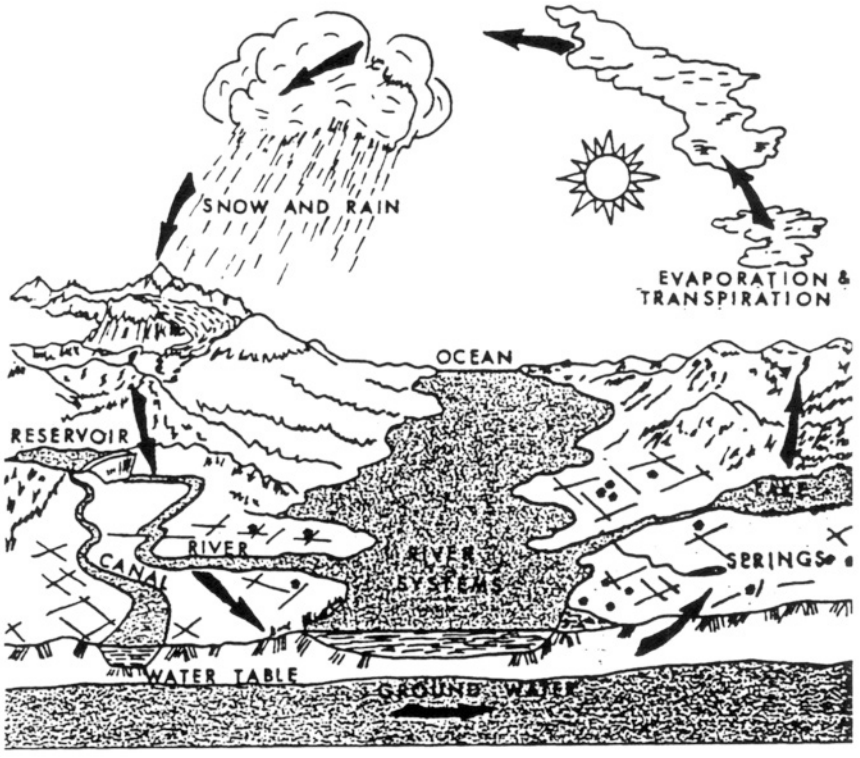


Figure 1.1—Nature's Waterwheel

The concept of water as a renewable resource stems from the hydrologic cycle. Hydrology, as the engineering science that analyzes the various components of this cycle, recognizes that the natural cycle can be altered by human and natural activities. For example, the long-term geologic forces that raise mountains can increase orographic precipitation on one side of the mountains and decrease precipitation on the other side with all of the attendant changes in stream flow, flooding, etc. In the short-term, the development and use of water modifies the natural circulatory pattern of the hydrologic cycle. For example, the use of surface water for irrigation may result in downward seepage from reservoirs, canals, ditches, and irrigated fields, adding to the ground water. Disposal of urban wastewater may recharge the ground water or change the flow in streams. All waters utilized in a non-consumptive manner may deteriorate in quality and create water quality problems in other portions of the hydrologic cycle. Diversions of stream flows impact downstream flows which, if transferred to other watersheds, impact the stream flows and ground water systems in the other watersheds. In addition, pumping from wells may reduce the flow of water from springs or seeps, increase the downward movement of water from the land surface and streams, reduce the amount of natural ground water discharge by evaporation and transpiration, induce the inflow of poorer quality water to the ground water reservoir, or have a combination of all these effects.

The natural circulation of the hydrologic cycle may be changed by actions not related to direct water use. Among these actions are weather modification activities (i.e., cloud seeding), drainage of swamps and lakes, waterproofing of the land surface by buildings and pavements, and major changes in vegetative cover (i.e., removal of forests and cultivation of additional agricultural lands).

The science of engineering hydrology attempts to account for and quantify all aspects of the hydrologic cycle. This handbook presents detailed engineering descriptions of the various phases of the hydrologic cycle.

Chapter 2

PRECIPITATION

I. INTRODUCTION

Precipitation is the primary source of the earth's water supplies. It includes all water that falls from the atmosphere to the earth's surface. Precipitation occurs in two forms that are of interest to hydrologists, liquid (rain and drizzle) and solid (snow, hail, and sleet).

Mean annual precipitation is a major climatological characteristic of a region, while other precipitation characteristics such as intensity, duration, and precipitation form (such as rain vs. snow) are attributes that are generally of most interest to practicing engineers and hydrologists. All these characteristics of precipitation are of interest because they determine the amount and timing of runoff and other hydrological concerns, such as the amount of available soil water.

This chapter deals with formation, types and distribution of precipitation, methods of measuring rainfall and snowfall, processing and interpreting precipitation records, frequency analysis, weather modification, and generation of synthetic weather records.

II. FORMATION AND TYPES OF PRECIPITATION

A. Mechanisms

Two atmospheric processes are primarily responsible for producing precipitation. These processes are generally referred to as: 1) the collision and coalescence process and 2) the ice-crystal process.

The collision and coalescence process is the means by which small water vapor droplets (created by condensation around available nuclei) increase in size in warm clouds. The term "warm" is relative, and refers to cloud conditions when temperatures are greater than 0° C. Larger cloud droplets descend more quickly toward the earth's surface under the pull of gravity than smaller droplets, and thus collisions occur as the larger droplets fall. As the water molecules in these droplets coalesce, the droplets become larger and fall more rapidly until they attain the size of raindrops (approximately 2 mm) and fall to the ground. The size of raindrops depends on several factors, including the vertical velocity (upward) supporting the droplets, the depth of the cloud, the time droplets remain suspended, and atmospheric temperature. The production of raindrops and, consequently, the precipitation that reaches the ground is a function of these factors plus the liquid water content of the cloud, the electric field in the cloud and of the droplets, and the relative size of the droplets.

Warm precipitation is usually restricted to tropical regions and some middle latitude storms in the summertime. The predominant precipitation-producing mechanism in middle and high latitudes is the ice-crystal process. In this case, clouds extend into the atmosphere above the 0° C level, and are thus known as cold clouds. Interestingly, even though ambient temperatures may be less than 0° C, liquid

water droplets can still exist and are known as supercooled droplets. Such droplets can exist in a liquid phase to temperatures as low as −40° C, below which all water is in the form of ice particles. Between −40° and 0° C a combination of ice particles and liquid water droplets usually exists.

The ice-crystal process is controlled by nucleation, or the formation of an ice embryo. Homogeneous nucleation refers to the formation of an embryo from a water droplet when no foreign material is present. In this case, the droplet freezes by the simple aggregation of water molecules. This process is possible only when temperatures are near or below −40° C, which is normally the case only in high, cirrus clouds and thus is not the controlling factor in producing precipitation.

In contrast, heterogeneous nucleation occurs when foreign matter is present upon which droplets can either form and aggregate, or with which droplets can collide and freeze spontaneously. The larger the nucleating material, the faster will be the process of freezing, which can occur at higher temperatures. For instance, droplets with diameters of 10 mm will freeze at −32° C but droplets with diameters of 10,000 μm will freeze at −15° C. In general, the higher, deeper and colder the cloud, the greater the probability that ice particles are present and are the controlling mechanism for precipitation production.

B. Types of Precipitation

As ice crystals fall, they may collide and stick to one another, creating snowflakes. If air temperatures are near or below 0° C throughout the lower atmosphere, precipitation recorded on the ground will most likely be in the form of snow. If melting occurs, precipitation at the ground will be rain.

Other precipitation types are possible. If a warm layer aloft is sufficiently deep, the snow will melt and become raindrops; however, if a layer of air near the ground is near or below 0° C the rain may re-freeze. If it freezes before hitting the ground, it has a very granular, icy structure and is known as sleet. If the cold layer near the ground surface is very shallow, the rain may not re-freeze before striking the surface but may freeze on contact with surface elements whose temperature are lower than 0° C (if the water droplets in the rain are supercooled). In this case, precipitation is called freezing rain. Depending upon the location in the United States, this event is called by various names, such as an ice storm, an ice glaze, a silver thaw, or a silver frost.

Rain is typically a name reserved for drops with diameters larger than 0.5 mm. For drops with diameters smaller than 0.5 mm, the common name is drizzle, which has a much lower fall speed, or terminal velocity, than larger raindrops.

In storms, the terminal velocity of water droplets is typically countered by upward moving air. If the vertical velocity of air is positive upwards at rates near or greater than a drop's terminal velocity, the drop will remain suspended. If the upward velocity suddenly decreases, as it can in convective storms, the drops will quickly fall to earth. A sequence of these events leads to the observation of showers on the ground. Typically, raindrops have diameters between 1 and 5 mm with respective terminal velocities of between 4 and 9 m/sec (Chow et al., 1988).

Snowflakes generally have much slower terminal velocities than raindrops. The structure of snow determines its terminal velocity, while the ambient temperature where ice crystals form determines snow structure. At temperatures just slightly below freezing, snow is generally in the form of thin plates which, if they encounter temperatures at or slightly above 0° C in their descent, can melt slightly and then aggregate to form large, mushy flakes. These flakes are extremely reflective to a radar beam, as will be discussed later in this chapter.

Sleet, or ice pellets, usually have diameters around 5 mm, are transparent, and bounce as they hit the ground. In some rare cases, particularly with deep cold air trapped in a valley or on the cold side of a stationary, precipitating warm front, sleet can accumulate to depths greater than 200 mm.

Freezing rain can be among the most damaging precipitation events that occur. If it persists for more than 12 hours at sufficient intensity and with surface temperatures below −2° C, ice accumulation can be extensive. Costs of damage caused by this form of precipitation can easily run into millions of dollars

through downed power lines, traffic accidents, and destruction of roofs of buildings. If the freezing droplets are smaller than 0.5 mm in diameter, the precipitation type is called freezing drizzle.

Another way in which ice particles can increase in size is by colliding with supercooled water droplets, a process known as riming. Pilots often contend with moderate to severe riming on planes when an abundance of supercooled water exists. In the atmosphere, this rimed ice will increase in size until buoyancy can no longer support it. If it is fairly small (between 2 and 5 mm) it is known as graupel, or snow pellets. These pellets are very firm and bounce like sleet on the ground surface.

In convective clouds, a graupel particle may be forced upward a number of times. If the updrafts are of sufficient strength and the quantity of supercooled water droplets is optimum, these graupel particles can grow to enormous sizes and are known as hail. Hailstones are typically no larger than 10 mm but, in severe thunderstorms with cloud tops well above the freezing level, hail has been measured with diameters exceeding 100 mm. The size of the stones is proportional to the strength and duration of the updrafts they encounter. Hail can be extremely devastating but typically falls over relatively small areas (1 to 10 km^2).

Snow grains are the remaining type of solid precipitation. These particles are smaller than graupel (less than 1 mm), are opaque and fall from stratus clouds. As such, their fall velocities are typically much less than graupel and they seldom accumulate to great depths at the land surface. In certain situations, snow grains can fall for long time periods.

C. Principal Causes of Precipitation

In middle and high latitudes, precipitation is typically the result of large scale weather systems. Large scale refers to systems with length scales usually larger than 500 km (also known as synoptic scale). Precipitation from such systems is seldom localized and amounts can be rather uniform over large areas.

Central to synoptic meteorology theory is the mid-latitude or polar-front cyclone model. First proposed by a group of Norwegian meteorologists around 1920, this model stresses the development of a low pressure center along a fairly straight, stationary line (Palmen and Newton, 1969). As the low pressure center deepens (drops in surface pressure), a wave forms with the center of the low at the apex of the fronts. Air rotates around the low cyclonically (counter-clockwise in the northern hemisphere), pulling cold air southward on the west side of the low and pushing warm air northward on the east side of the low (in the northern hemisphere). The lines of demarcation are known as fronts, which often are regions of fairly distinct changes in weather (temperature, humidity, cloudiness, wind direction, precipitation, etc.). Major precipitation areas associated with this classical cyclone model are to the north of the warm front, extending cyclonically around the low center northward and westward some distance. This area of precipitation is often rather uniform, with maximum amounts typically 50 to 250 km north of the low pressure center. Another area of precipitation in this model is along the cold front where cold air rushing in at the surface causes warmer air ahead of the front to ascend, often producing convective types of clouds. This precipitation area is typically more broken and localized, with heavy showers and thunderstorms along and immediately behind the front.

Precipitation produced by mid-latitude cyclones is a function of the quantity of water available in the atmosphere and the strength of the dynamic processes which create the clouds and vertical motions around the low. It is now known that jet streams, which are regions of high velocity air some hundreds to thousands of meters above the ground, have a controlling influence on the development and movement of mid-latitude cyclones. Rather than continuous channels of fast air, jet streams typically are cores of high velocity movement through which air parcels travel. Jet streams are the result of large thermal contrasts in the atmosphere. The polar jet usually divides very cold polar air from milder, mid-latitude air; while the subtropical jet divides very warm, moisture-laden tropical air from cooler and (usually) drier mid-latitude air. At certain times, both of these jet streams can be quite active, with their energy generating significant cyclonic storm activity from latitudes as low as 25 or 30 degrees poleward to 80 degrees. Numerical models of mid-latitude synoptic weather are very sensitive to the strength and

position of these jet streams, and the transport of jet stream energy to the surface is critical in determining the fate of low pressure systems and the duration, coverage, and intensity of precipitation associated with them.

Convective precipitation can be associated with mid-latitude cold fronts but can also develop in certain situations when no synoptic scale system is active. Air mass showers and thunderstorms are examples of this type of convective activity and are most prominent in mid-latitude summer. Triggering mechanisms for this type of precipitation include upper air systems, which fail to develop a well-defined low pressure center, and very strong daytime heating in the presence of sufficient atmospheric water vapor. This latter situation can lead to extremely heavy rain and severe thunderstorms if the atmosphere is hydrostatically unstable; that is, it is basically stable in the lower boundary layer and somewhat unstable above but becomes very unstable if surface heating and other forces are sufficient to create vertical motions strong enough to penetrate through this layer. This is a potentially explosive atmosphere and is a frequent cause of severe convection and heavy precipitation developing in very short (30–60 min) time periods.

A third element contributing to precipitation development is orography, or over-mountain impact on air flow. As air flows over mountains it is forced upward. This vertical displacement of air is somewhat analogous to warm air being forced upward in the presence of a cold front. As air ascends, it expands and cools. This is a principle directly taken from the ideal gas law, sometimes referred to as the equation of state:

$$pV = mRT \tag{2.1}$$

where p is air pressure, V is volume, m is mass, R is the gas constant for 1 kg of air, and T is temperature (K) (Wallace and Hobbs, 1977). Since m/V is simply air density (ρ), this reduces to:

$$p = \rho RT \tag{2.2}$$

Both air pressure and air density decrease from the surface upward, and thus air parcels forced upward are moved into an environment of lower pressure and density. This causes the parcels to expand and cool, since temperature is indicative of molecular activity. Thus, air forced over a mountain by the wind will cool as it rises. As it cools, it may reach a level where temperature and dewpoint are equal if the absolute water vapor content of the air parcel does not change. This level is referred to as the lifting condensation level (LCL), because water vapor begins to condense into liquid water droplets and the relative humidity is around 100%. At this level, clouds form and if vertical motions are strong enough, precipitation can develop. Even on relatively clear days, a combination of daytime surface heating and wind flow leads to cloud development over mountain ridges.

If a cyclonic system is already producing clouds and precipitation, the air flow over mountains will enhance precipitation on the windward slope. Thus, it is common for precipitation to increase with elevation in mountainous areas. Because precipitation-elevation relationships change with region and with storm events, simple approaches to estimate these relationships are not necessarily reliable. Precipitation is strongly controlled by elevation in the western United States, where mountain snowpacks supply most of the water for the region.

Tropical storm systems can produce significant amounts of precipitation and cover relatively large surface areas. Such systems typically affect the United States between July and November, with peak activity in the late summer. Atlantic tropical systems originate as tropical waves off the west coast of Africa and move toward the Caribbean on the predominate easterly flow across the ocean. Some systems develop into hurricanes, which means sustained winds are in excess of 33 m/sec (74 mi/hr). The path taken by tropical storms is usually dictated by the winds aloft, which are generally weak easterly winds in this region. Strong upper level winds associated with jet streams are actually destructive to tropical

storms. Larger hurricanes can sometimes have radii in excess of 500 km and their movement is much less directed by winds aloft, thus making prediction of their movement difficult.

Tropical cyclones are known as "warm core lows" because they are warmer than their surroundings from the surface upward. They are strongest at the surface, and winds and pressure gradients diminish with height. In contrast, mid-latitude or extratropical cyclones have colder temperatures in the storm center than around them, and the intensity of the low increases with increasing height in the atmosphere, thus making it a so-called "cold core low." Tropical cyclones usually have cloud heights above the freezing level, so that even in these warm systems there is typically a mixture of warm and cold precipitation-producing mechanisms in place.

The quantity of precipitation that falls from tropical systems (as well as extratropical systems) is a function of storm movement, relative location to the storm center, and storm movement relative to land masses. Even relatively weak tropical cyclones have sometimes produced extremely heavy rainfall, especially when the storm stalls over an area or moves perpendicularly into a mountain ridge. Central pressure of a hurricane is a good gage for maximum wind speeds but is a poor indicator of total precipitation. In general, the lower the pressure in the center of the storm, the more energy the storm possesses with stronger horizontal and vertical wind speeds. This condition also creates more potential for heavy precipitation; however, strong storms sometimes have relatively fast forward speeds which prevent large total accumulations, although short-duration intensities can be quite impressive.

Precipitation around a tropical cyclone, particularly one which has become fairly well-organized and concentric, usually comes from rain bands rotating cyclonically around the low pressure center. These bands of showers and thunderstorms typically increase toward the center of the storm (or "eye" in a well organized hurricane), reaching the maximum intensity in the eye wall, where a solid circle of severe thunderstorms is usually located. Often the heaviest precipitation is in these "eye wall thunderstorms" and in rain bands that are generally to the east of the cyclone center.

In the southeastern United States, tropical storms are responsible for as much as 5 to 30% of the normal precipitation in the summertime. For the remainder of the nation, the numbers are much smaller, although tropical cyclone activity in the eastern Pacific Ocean sometimes produces heavy precipitation in the Southwest. In some cases, desert locations in Arizona and southern California can receive most of their annual precipitation from the remnants of these storms. These regions have a monsoonal climate dominated by notable wet and dry seasons. Tropical storms are sometimes responsible for ending droughts, especially in the Southeast, and can actually have beneficial results.

III. VARIATIONS IN PRECIPITATION

A. Geographic Distribution

1. Latitudinal Variations Globally, average annual precipitation (Figure 2.1) is generally heaviest in tropical regions and decreases poleward, indicative of the diminishing capacity of air to hold moisture with decreasing temperature; however, there are significant deviations from this mean trend. Latitudes near 30 degrees have relatively little precipitation because of the climatic propensity for air to rise near the equator and then descend near these latitudes. Some of the world's great deserts (the Sahara, the Middle East, the Australian interior, the southwest portion of Africa, and the American Southwest) are in this region, showing that this sinking air is counter-productive to precipitation production.

Poleward moving air typically rises again in middle latitudes, reaching an average maximum at about 60 degrees latitude. Precipitation enhancement occurs in these regions, with more frequent cyclonic activity.

In addition to the cellular structure of poleward moving air (Figure 2.2), other dominant forces shaping regional precipitation are the general circulation of both the oceans and the atmosphere and their relationship to the shape and position of continents. In general, flow in the major oceans is cyclonic,

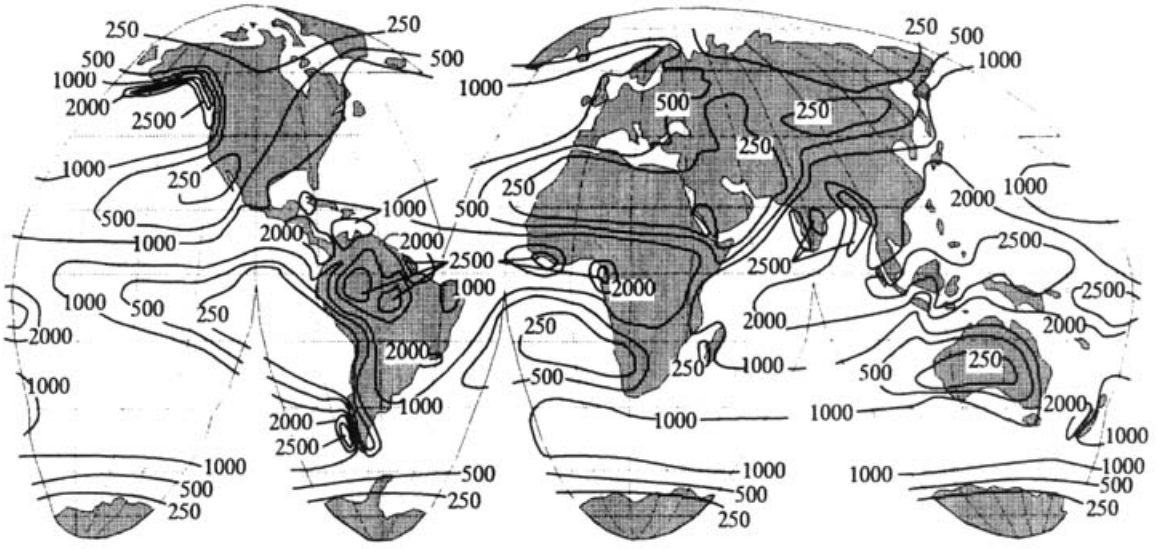


Figure 2.1.—Average Annual World Precipitation (mm)

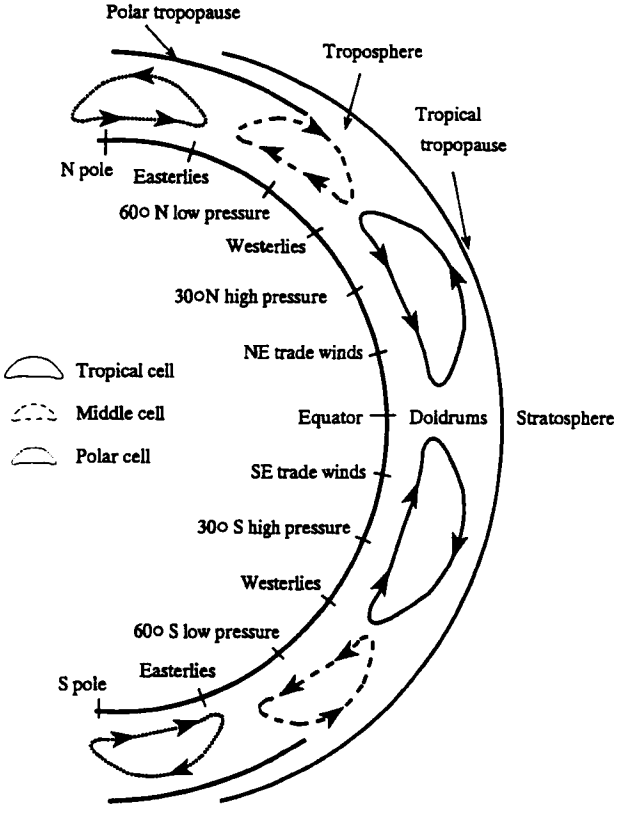


Figure 2.2.—Three Cell Atmospheric Circulation Patterns for Both Hemispheres

which means tropical water and atmospheric moisture are transported poleward along the east sides of continents. Therefore, eastern portions of North America, Australia, and the EurAsian continent have relatively more precipitation than their counterparts on the west coast, especially during their respective summers. Slight exceptions to this phenomenon do occur, such as along the north Pacific coast of the United States and Canada where a persistently strong jet stream in the winter and the circulation around

the semi-permanent Aleutian low brings copious precipitation to these regions, which is enhanced by the sharp increase in elevation immediately inland from the coast.

Polar regions are typically under the influence of sinking air, plus the very cold atmosphere can hold relatively little water vapor. As a result, these areas have very little precipitation which for most if not all of the year, is locked in the solid forms of snow and ice.

2. ***Distance from a Moisture Source*** Perhaps the second most important factor in determining precipitation at a given location is its distance from a moisture source—i.e., an ocean. Continental interiors typically have less precipitation because of a minimum of precipitable water and a lack of larger, ocean derived salt particles, which are better nucleating materials than dust and other materials from land (Figure 2.3).

It is estimated that only 15% of the nearly $5.68 \times 10^{15} m^3$ of water that evaporate globally each year come from continental areas; the other 85% is evaporated from the oceans (Ahrens, 1991). The Amazon basin in South America is a notable exception to this rule. Two contributing factors to the heavy rainfall in this region are its abundant reserve of water held in the jungle and its equatorial location, where horizontal movement of storm systems is minimal. Thus, almost daily rains occur in this region and in similar interior sections of equatorial Africa.

To a lesser extent, water bodies smaller than oceans can exert influence on precipitation. The Great Lakes, for instance, are responsible for enhanced precipitation, especially in the wintertime when strong cold air advection over the relatively warm water produces significant snow squalls on windward shores. "Snow belts" are well-known in this region, but the exact location of maximum snowfall and the amount that falls is dictated by the wind direction relative to the water, the fetch over the water, the temperature contrast between the water and the advected air, and the elevation rise on the windward shore. In contrast, these same lakes can actually be responsible for precipitation reduction in the warmer months of the year because the water is relatively cooler than the land at this time, creating an air circulation pattern away from the lakes. The Great Lakes region then has a relatively higher surface pressure than the land area around it, leading to subsidence, or sinking air aloft, which is detrimental to cloud and precipitation production.

Sea breeze circulations are another example of localized wind flows created by the contrast in temperature and moisture between land and water. These diurnal circulations are dominated by an ocean to land wind flow (very regular in the tropics; pronounced in the summertime in mid-latitudes). Warm, moist air flowing inland in the Gulf Coast and Atlantic Coast regions of the United States, for instance, typically converge at the surface with warmer and drier air moving oceanward each afternoon. Surface convergence leads to upward vertical motion and, especially in the presence of daytime heating, cloud development. Precipitation is thus enhanced in regions immediately inland from coastal sections in the summertime, usually at distances from 10 to 50 km from the shoreline. This scenario is particularly apparent in the Florida peninsula where, for much of the year, these circulations create thunderstorms over the land each afternoon while satellite images reveal very little cloud development or precipitation over the nearby ocean.

3. ***Orographic Influences*** As mentioned in the section on principal causes of precipitation, precipitation is normally enhanced in the vicinity of mountains. Enhancement is dependent on several factors, including wind direction (relative to topography), wind speed, atmospheric moisture (precipitable water), elevation rise, and slope angle. For these reasons, orographic precipitation is most pronounced during the winter months in middle latitudes when atmospheric flow is strongest; however, convective precipitation in summer months also is enhanced over mountains due to diurnal winds which tend to move up slopes and through valleys during the day, and reverse their direction at night (Whiteman, 1990).

Orographic precipitation produces marked contrasts in seasonal precipitation distribution, which must be considered for any type of system design in mountainous areas. The monthly distribution of precipitation for two sites only 10 km apart in the Reynolds Creek Watershed in Idaho (Fig. 2.4) is

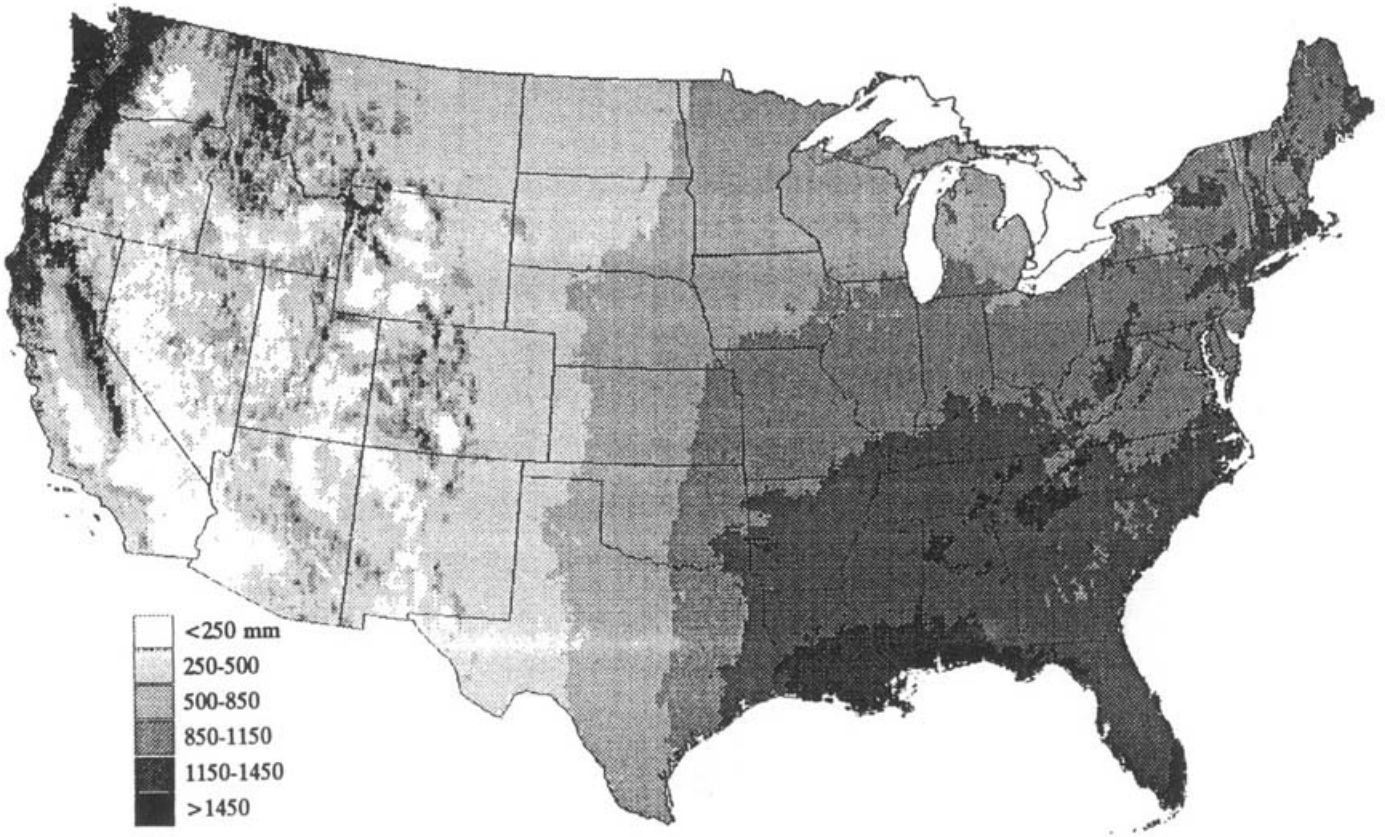


Figure 2.3.—Long-term Annual Precipitation for the Continental United States, Data from NWS/Cooperator and SCS Snotel Stations having at Least 20 yr of Valid Measurements. Map Produced using the PRISM Model, a New Statistical/Topographic Precipitation Mapping Scheme (Daly et al., 1992). For further information contact Oregon State

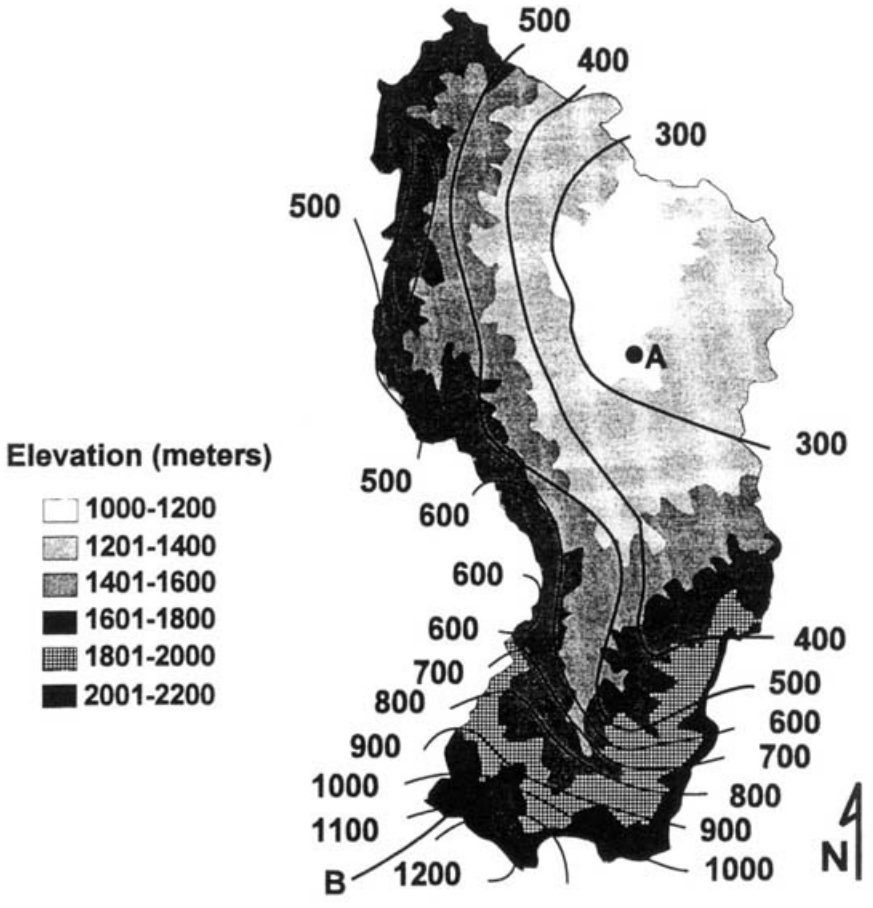


Figure 2.4.—Reynolds Creek Watershed, Idaho; Elevation Contours are Shaded, with Overlaying Isohyets of Mean Annual Precipitation (mm)

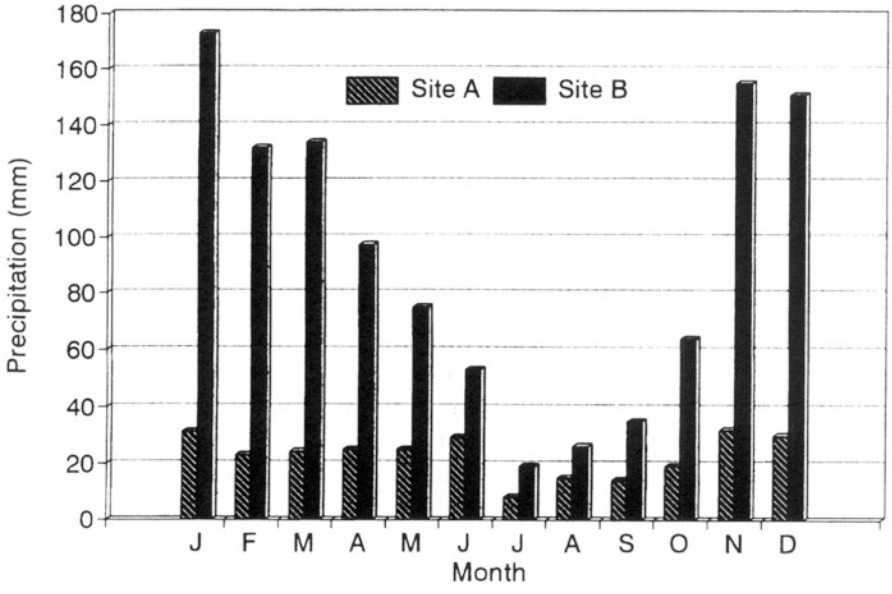


Figure 2.5.—Monthly Precipitation Distribution (mm) at sites A and B, Reynolds Creek Watershed, Idaho

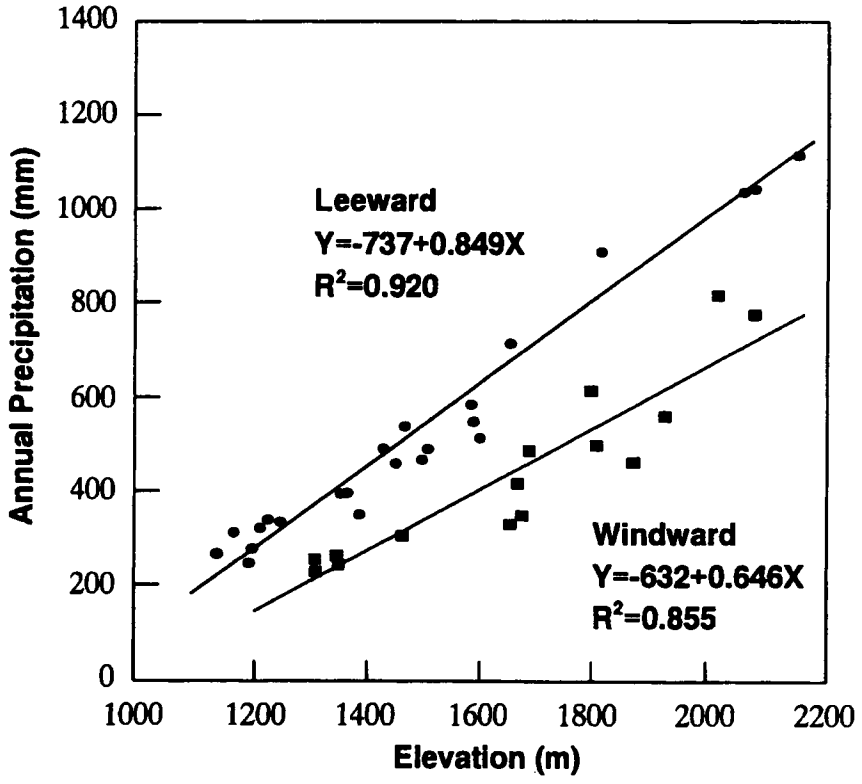


Figure 2.6.—Annual Precipitation (mm) vs. Elevation (m) for Windward and Leeward Locations, Reynolds Creek Watershed, Idaho

indicative of many mountainous areas and is shown in Figure 2.5 (Hanson and Johnson, 1993). Site A has an elevation of 1193 m with an annual precipitation of 275 mm (22% as snow) and maximum monthly precipitation in June. In contrast, site B, at an elevation of 2164 m, has an average annual precipitation of 1114 mm (76% as snow), and monthly precipitation is greatest in January. This watershed (at 43 degrees N latitude) is subject to frequent winter storms, usually accompanied by strong winds. The low elevation site (A) is downwind of site B and is in a relative precipitation "shadow." Such shadows are common on the lee-side of mountain ranges (Sumner, 1988). This is well illustrated in the map of annual precipitation for the United States (Fig. 2.3), and discussed by Daly and Neilson (1992). Where atmospheric flow is predominantly from one direction, it can lead to climatic dry zones in areas such as east of the Sierra Nevada and Cascade mountains in the western United States and even in relatively humid locations such as around Asheville, North Carolina, which is in a shadow region of the southern Appalachians.

As mentioned in the last section, precipitation-elevation relationships can sometimes be different from one side of a mountain range to the other. For example, on the Reynolds Creek Watershed, the best-fit regression line of precipitation with elevation is different for the windward and the leeward locations (Figure 2.6).

Design applications often require information about the relationship between a standard time increment for precipitation and expected precipitation in a different time interval. It is important to realize that such relationships change with elevation and other topographic factors in mountainous areas. As an example, the ratio between 1-hour precipitation and precipitation for shorter and longer durations was calculated for sites A and B at Reynolds Creek Watershed, for all return periods (Table 2.1). At the low elevation site (A) the ratio is just 1.83, while at high elevation site B the ratio is 5.01—nearly three times greater. Thus, if only 24-hour precipitation were available (as is often the case), estimated 1-hour precipitation at site A would be about half as much as the measured 24-hour amount; at site B, however,

TABLE 2.1. Ratio Between 1-hr Precipitation (mm) and Precipitation for n-min Durations at Sites A and B, Reynolds Creek Watershed, Idaho

	Site	
Duration	A	B
5 min	.45	.43
10 min	.67	.66
15 min	.77	.78
20 min	.79	.80
30 min	.87	.92
2 hr	1.09	1.20
6 hr	1.40	2.01
12 hr	1.61	3.36
24 hr	1.83	5.01

the 1-hour amount would be only one-fifth as much as the 24-hour amount. Such a difference could be crucially important in engineering design at the two sites, and such considerations should be made in estimating precipitation in mountainous regions.

B. Time Variation

Temporal precipitation variation on longer time scales is largely driven by normal fluctuations in atmospheric flow with known periodicities. For example, monsoonal climates are characterized by pronounced wet and dry seasons. Arizona and New Mexico are examples of this type of climate, with a relatively wet season usually beginning in June and continuing through the summer. Other times of the year are normally very dry, except for some precipitation increase in the winter if the subtropical jet stream is strong enough, and storm systems are actively moving east across northern Mexico and California.

Meteorologists are becoming increasingly aware of atmospheric teleconnections—that is, perturbations in the oceans or atmosphere that have an effect on the weather in locations downstream. While it may be extremely difficult to know the cause of perturbations, an understanding of the teleconnections associated with them may lead to greatly improved long-range weather forecasts.

A well-known example among these perturbations is the El Niño-Southern Oscillation (ENSO) phenomenon in the tropical Pacific Ocean. In some years, the normal east to west flow of air and water in the equatorial Pacific is reversed, increasing temperatures and precipitation in the far eastern Pacific (especially along coastal Ecuador and Peru). El Niño greatly modifies the entire energy balance in the Pacific, which is a huge energy pool for the world. As a result, changes in the weather and ocean water in the Pacific result in somewhat predictable changes in other parts of the world. Globally, some regions are positively correlated with El Niño while others are negatively correlated (Figure 2.7). In the United States, highest positive correlations are associated with winter precipitation in the Gulf Coast States and the Southwest, and with summer precipitation in the southern Rockies (Ropelewski and Halpert, 1987; Woolhiser et al., 1993). In contrast, the southern oscillation index (SOI, a measure of ENSO in which negative values indicate the presence of an El Niño event) was found to be negatively correlated with October through March precipitation in the Pacific Northwest, especially in mountain areas (Redmond and Koch, 1991). In other words, during an El Niño event, the Northwest is normally dry while the Southwest is often wet (correlations around 0.5 for both). This is but one example of atmospheric teleconnections that can directly influence temporal precipitation variability.

The seasonal distribution of precipitation in an area is an important consideration for many reasons. For example, growing season precipitation dictates the need for irrigation. In the eastern United States, precipitation is normally adequate during the growing season and, thus, irrigation usage is relatively low.

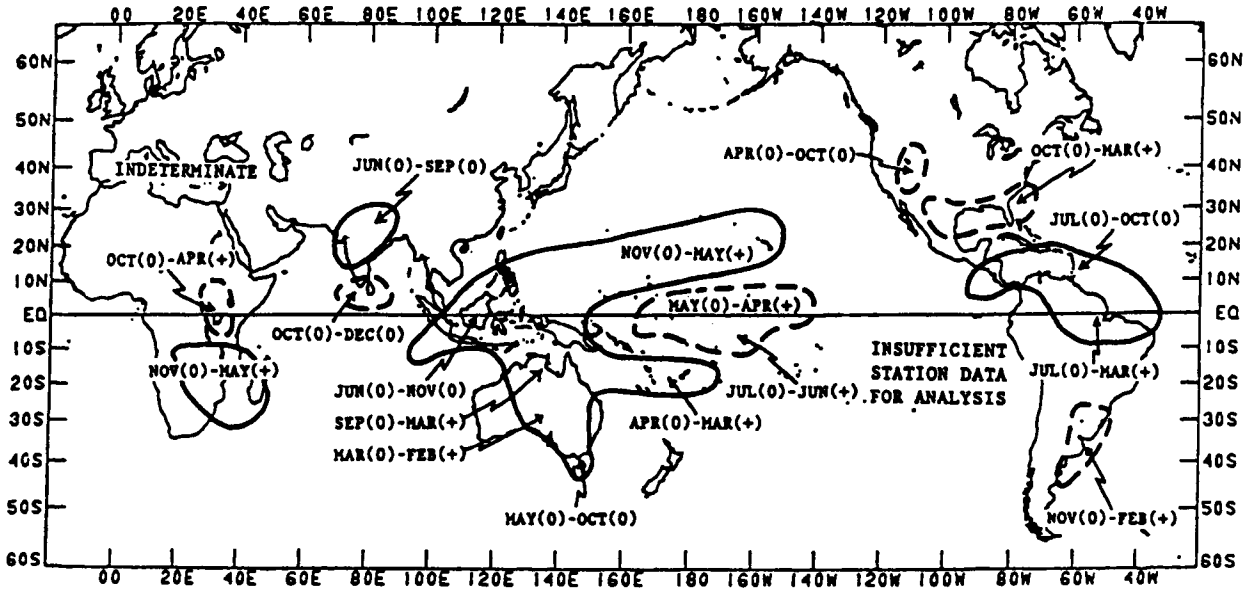


Figure 2.7.—Regional Enhancements (dashed lines) and Diminutions (solid lines) of Precipitation During El Niño Episodes. [Months of concentration are shown for each region. Year of associated anomalously high sea surface temperatures in the tropical Pacific is noted by month of concentration (0 = same year; + = subsequent year)] (from Ropelewski and Halpert, Monthly Weather Review, Vol. 115, p. 1625, American Meteorological Society, 1987)

In the western United States, irrigation is largely responsible for agricultural production. Because irrigation water supplies are normally dependent on winter snowpacks, water for streamflow and reservoir recharge are of great concern. The very dry summer seasons in the West make winter precipitation crucial to survival, while in the East, where frequent precipitation is normal, water supply systems are more vulnerable to relatively short-term droughts that occur especially during the warm season, when both water requirements and evapotranspiration are high.

Monthly precipitation distributions for selected locations in the United States are shown in Figure 2.8. Regional characteristics are quite evident. From the Northeast to the Central Gulf Coast, precipitation is nearly uniform for all months. In the Southeast, a summer to early fall precipitation maximum is noted. Very little winter precipitation with a summer maximum is characteristic of the Plains and Midwestern states. Low elevation locations in the Inter-mountain region have uniformly low precipitation year-long, while higher elevations have precipitation distributions resembling the graph in Figure 2.5. The Pacific Coast region is characterized by wet winters and extremely dry summers—exactly opposite to the Southeastern states.

On shorter time scales, precipitation variability is a result of 6-hour to 3-day synoptic events (frontal passages, cyclonic storms) and/or diurnal fluctuations, chiefly resulting from the daily cycle of insolation. Synoptic-scale events are driven by the wind flow at upper levels of the troposphere. The persistence of synoptic patterns can lead to prolonged periods of dry or wet conditions.

Diurnal precipitation variability is greatest during the warm season and is thermally driven. Thus, convective precipitation predominates during the summer and is induced by simple surface heating or in combination with thermally-forced flows, such as sea breezes or mountain/valley winds. It is also a function of available water. Therefore, the highest frequency of diurnal precipitation is found in the southeastern United States where all of these ingredients are maximized. Thunderstorms comprise a large portion of the precipitation events on these relatively short time scales, and a map of the annual number of thunderstorm days each year in the United States is shown in Figure 2.9. Obviously, thunderstorm enhancement in the eastern United States is partially due to synoptic-scale storms, but a large

portion is simply due to mechanisms operating on daily time scales. Note the maximum over the interior of the Florida peninsula, a region frequently experiencing sea breeze convergence in the summer, when heat and moisture are greatest.

C. Extreme Precipitation Events

The maximum quantity of precipitation that can be expected at a location in a given period of time is related to atmospheric water vapor content (called precipitable water) and the strength of vertical motion. Precipitable water is defined to be the total mass of water in a column of atmosphere, given by:

$$m_p = \int_{z_1}^{z_2} q_v \rho_\alpha A dz \tag{2.3}$$

where m_p is the total precipitable water, Z_1 and Z_2 are elevations, q_v is specific humidity, ρ_a is air density, and A is the cross-sectional area (Chow et al., 1988). Southeastern sections of the United States have a much greater amount of precipitation possible in any given year because precipitable water values are highest there and maximum vertical motions are typically much greater than locations further north and west. Vertical motion is a function of lifting forces related to conditional instability, tropical storms, jet stream dynamics, etc. This is not to say that vertical motions in northwestern sections of the nation for example are not sometimes quite strong; however, they typically occur in winter when precipitable water values are much lower.

The World Meteorological Organization (WMO) states the following equation as an approximation for estimating the world's extreme precipitation values:

$$P = 422 T_d^{0.475} \tag{2.4}$$

where P is the precipitation depth in millimeters and T_d is the duration of the event in hours (WMO, 1983). A listing of the world's greatest recorded precipitation amounts is given in Table 2.2. These, of course, represent the current upper bounds on recorded precipitation. Most locations will never come close to receiving these extreme amounts.

Maximum expected precipitation is dependent on both the time of year and elevation. Accompanying graphs of average winter, summer, and annual precipitation (Figure 2.10), and 2-year, 6-hour precipitation versus elevation (Figure 2.11) over the Reynolds Creek Watershed show significant differences between low and high elevation sites. Other climatic regions would most likely have slightly different graphs, but seasonal and elevational dependence would be noted in most locations. These concepts are combined to calculate probable maximum precipitation (PMP). A more complete treatment of PMP is given in the section on Processing and Interpreting Precipitation Records in this chapter and in Chapter 8, Floods.

IV. THE MEASUREMENT OF PRECIPITATION

The measurement of precipitation dates back at least to the 4th century B.C., when a network of rain gages was established in India (Biswas, 1967). Rain gages were used in Palestine in the 1st century B.C., in China in the 13th century A.D., and in Korea in the 15th century (Biswas, 1970). The shape of the gages in India is not known, but the Chinese and Korean gages were cylindrical or barrel-shaped. The Korean gages were about 30 cm deep with 15 cm diameter cylinders, and had about the same characteristics (and accuracy) of many of the rain gages in widespread use today. Rain gages were first used in Europe in the

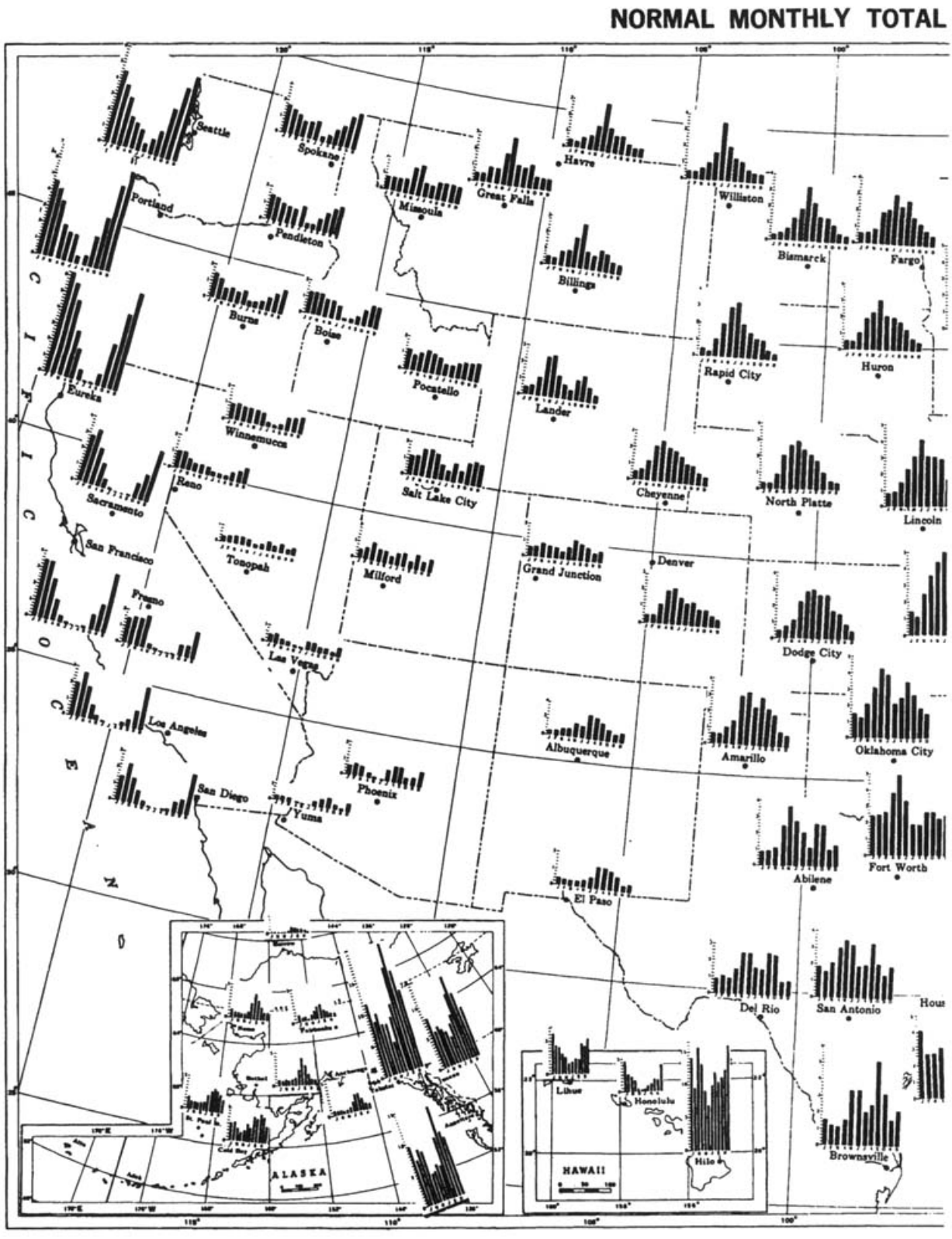
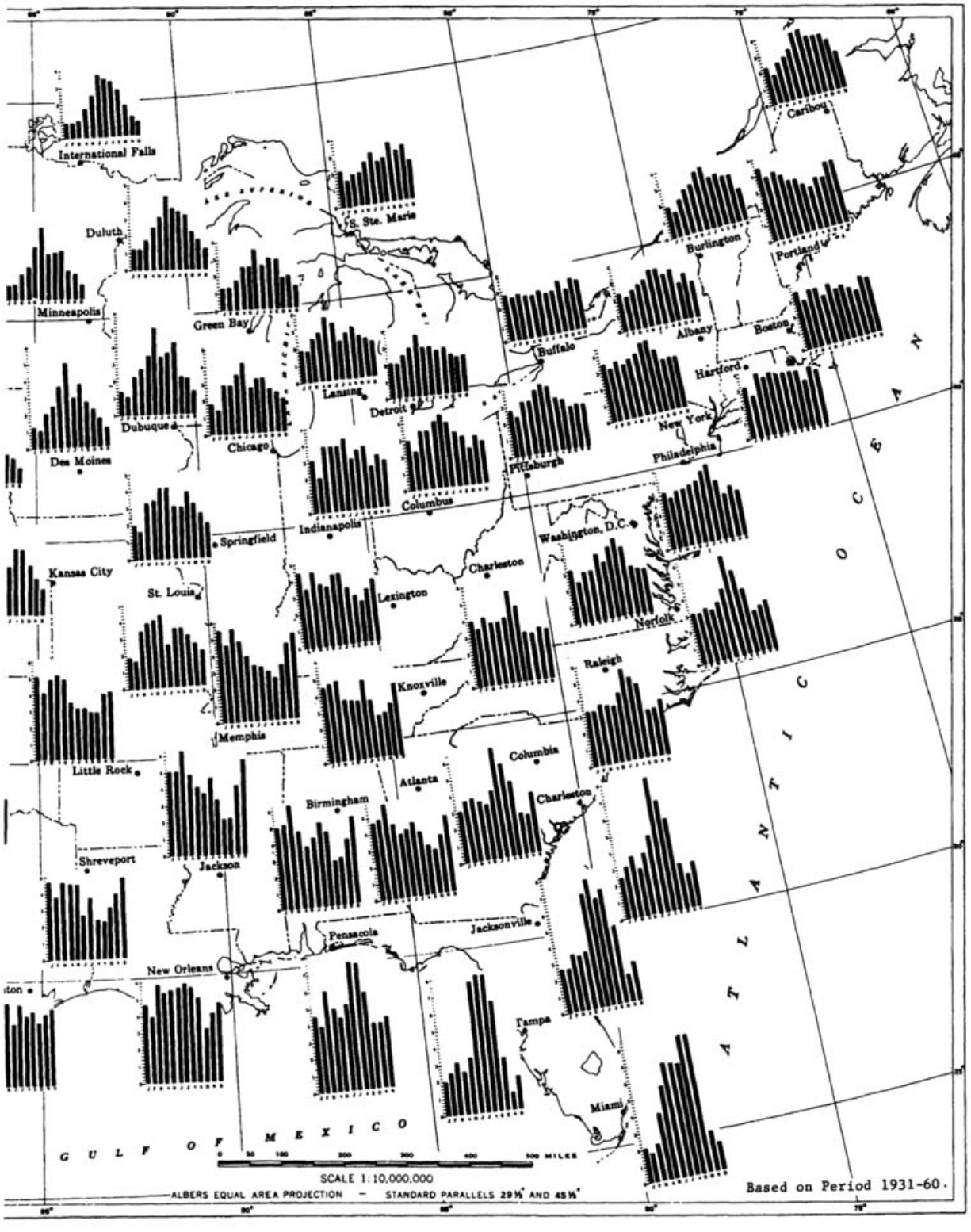


Figure 2.8.—Monthly Precipitation Distribution in the United States (1 in = 25.4 mm; from Climatic Atlas of the United States, U.S. Dept. of Commerce, 1979)

PRECIPITATION (Inches)



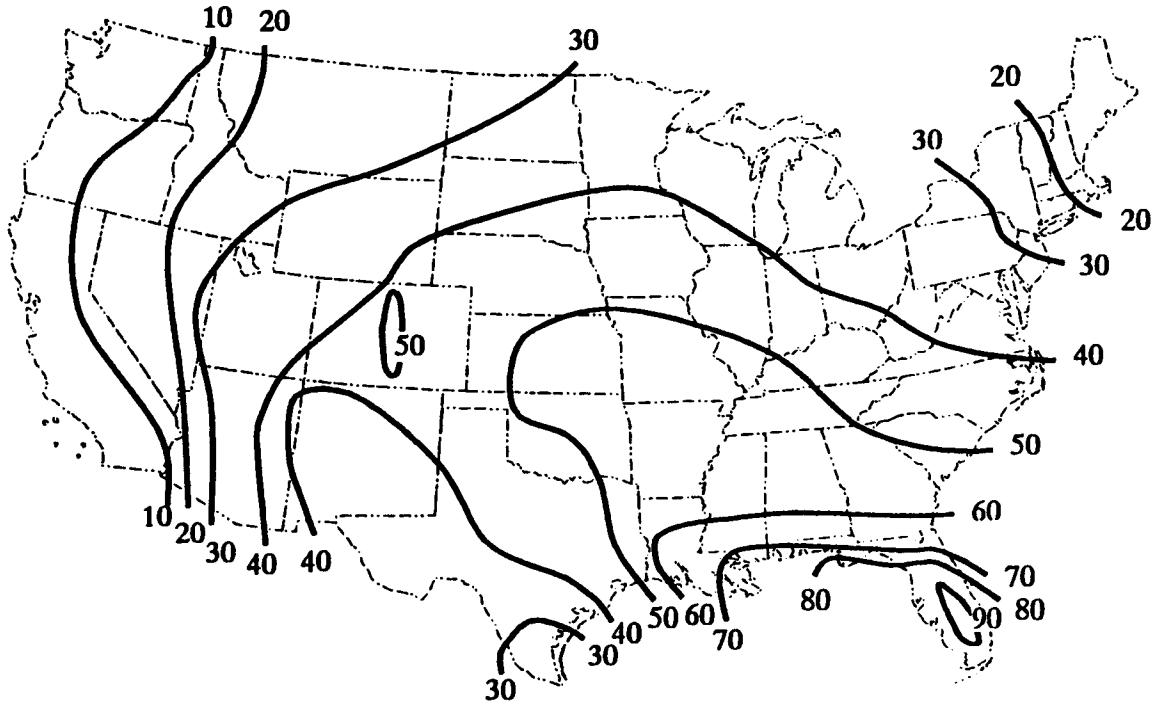


Figure 2.9.—Average Annual Thunderstorm Days for the United States

TABLE 2.2. Precipitation Extrema for the World (from Griffiths 1985, p. 125. In: David O. Houghton, 1985, Handbook of Applied Meteorology, John Wiley & Sons, Inc., New York, NY. Reprinted by permission of John Wiley & Sons, Inc.)

Parameter of Precipitation	Precipitation (in.)	Precipitation (mm)	Station
Highest annual total	1042.	26470.	Cherrapunji, India (Aug. 1 1860–July 31, 1861)
Highest annual means	460.	11680.	Mt. Waialcale, Kauai, Hawaii (32 yr)
	450.	11430.	Cherrapunji, India (74 yr)
	405.	10290.	Debundscha, Cameroons (32 yr)
Lowest annual mean	0.02	0.50	Wadi Halfa, Sudan (39 yr)
	0.03	0.75	Arica, Chile (59 yr)
Highest monthly total	366.	9300.	Cherranpunji, India (July 1861)
Highest monthly mean	106.	2700.	Cherranpunji, India (July 1861)
Highest 5-day total	150.	3810.	Cherrapunji, India (Aug. 1841)
	115.	2920.	Jamaica (Nov. 1909)
Highest 1-day total	73.62	1870.	Cilaos, La Reunion (Mar. 16, 1952)
Highest 12-hr total	52.76	1340	Belouve, La Reunion (Feb. 28–29, 1964)
Highest 4.5-hr total	31.00	787.	Smethport, Pa. (July 18, 1942)
Highest 42-min total	12.00	305.	Holt, Mo. (June 22, 1947)
Highest 20-min total	8.10	205.	Curtea-de Arges, Rumania (July 7, 1889)
Highest 1-min total	1.23	31.2	Unionville, Md. (July 4, 1956)
Highest no. of rain days in 1 yr			Cedral, Costa Rica (1968) - 355 days
			Cedral, Costa Rica (1967) - 350 days
			Behia Felix, Chile (1916) - 348 days
Highest annual total (snow)	1017.	25830	Paradise Ranger Station, Mt. Rainier, Wash. (1970–1971)
Highest annual means (snow)	582.	14780.	Mt. Rainier, Wash.
	575.	14600.	Crater Lake, Ore.
Greatest depth on ground	454.	11530.	Tamarack, Calif. (Mar. 9, 1911)
Highest monthly total (snow)	390.	9900.	Tamarack, Calif. (Feb. 1911)
Highest 12-day total (snow)	304.	7720.	Norden Summit, Calif. (Feb. 1–12, 1938)
Highest 6-day total (snow, single snowstorm)	174.	4420.	Thompson Pass, Alaska (Dec. 26–31, 1955)
Highest daily total (snow)	76.	1930.	Silver Lake, Colo. (April 14–15, 192

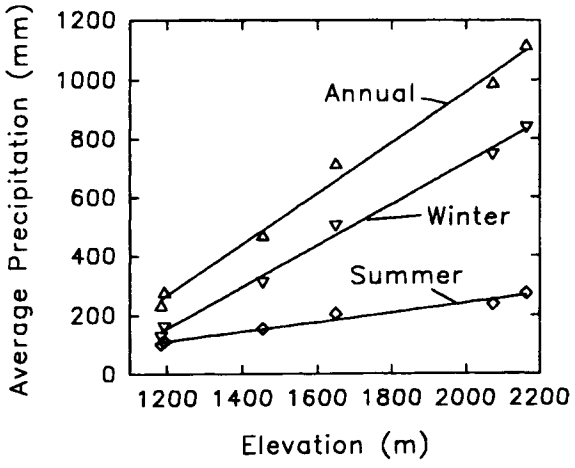


Figure 2.10.—Average Seasonal (winter and summer) and Annual Precipitation (mm) vs. Elevation (m) for Six Sites on the Reynolds Creek Watershed, Idaho

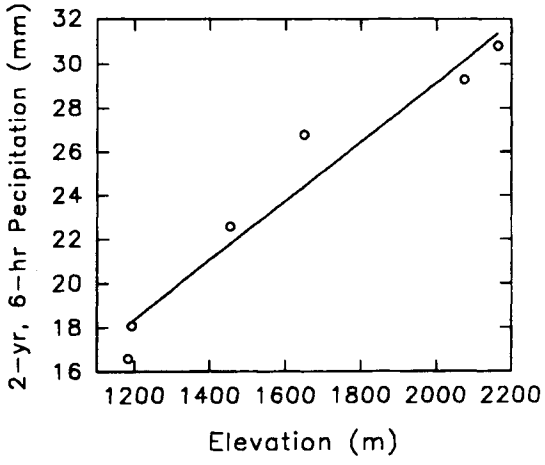


Figure 2.11.—2-yr 6-hr Precipitation (mm) vs. Elevation (m) for Six Sites on the Reynolds Creek Watershed, Idaho

17th century, including a tipping-bucket gage developed by Sir Christopher Wren and modified by Robert Hooke in 1678 (Biswas, 1970).

The 18th century was marked by the development and use of numerous designs of gages around the world. In 1802, Dalton stated (cited in Biswas, 1970):

> The rain gauge is a vessel placed to receive the falling rain, with a view to ascertain the exact quantity that falls upon a given horizontal surface at the place. A strong funnel, made of sheet iron, tinned and painted, with a perpendicular rim two or three inches high, fixed horizontally in a convenient frame with a bottle under it to receive the rain, is all the instrument required.

The measurement of the "exact quantity" of rain (or snow) that falls upon a horizontal surface has been the subject of a large number of investigations in the past two hundred years. As cited in Rodda (1967), Heberden (1769) established that elevated gages caught less rainfall than lower gages. This was shown conclusively by Jevons (1861) to be due to wind action (cited in Nipher, 1878). To eliminate the effects of wind on the collection of rainfall, Stevenson (1842) developed a gage with an orifice at ground level (cited in Rodda, 1967). This ground-level gage, also known as a pit gage, properly sited and protected against splash, is perhaps the most accurate gage for rain. A ground-level gage has been used by the WMO as a reference for the various national gages (Rodda, 1970). Ground-level gages are not suited for large networks, however, due to much the same objections listed in Nipher (1878). The pit gages fill with leaves or snow due to drifting and blowing, are prone to damage, and to float with accumulated water around the gage, and produce large measurement error with small differences in orifice height. The problems with snow measurement led Nipher to design the first shielded snow gage; a design that is still in use.

In the past 100 years, several types of gages for the measurement of precipitation have been developed. New techniques have been devised to measure precipitation with both direct sensing and remote sensing. In spite of all the latest technology, most precipitation measurements in the United States are still made with a measuring stick using a variation of the type of gage described by Dalton in 1802. For many types of rainfall or snowfall measurements, the "exact quantity that falls on a horizontal surface" must be known with a high degree of accuracy. Overall, a perfect, cost-effective system for precipitation measurement has yet to be demonstrated.

Chapter 3

INFILTRATION

I. INTRODUCTION

Infiltration is commonly defined as the process of water entry at the land surface into a soil from a source such as rainfall, irrigation, or snowmelt. The rate of infiltration is generally controlled by the rate of soil water movement below the surface. Rainfall excess is the portion of applied water that leaves the surface site not as infiltration but as runoff. As can be seen in Figure 3.1, the classic components of the hydrologic cycle for an event are 1) evaporation, 2) interception and depression storage, 3) infiltration, and 4) rainfall excess. The difference between rainfall excess and infiltration models is that rainfall excess models lump the losses (infiltration, evaporation, interception, and depression storage) together while infiltration models only describe the infiltration portion. Since evaporation, interception, and depression storage are normally minor compared to the infiltration portion during an event, rainfall excess models can be considered synonymous with infiltration models. This chapter of the handbook presents current knowledge and practice of modeling infiltration and rainfall excess.

II. PRINCIPLES OF INFILTRATION

The time-dependent rate of infiltration into a soil is governed by the Richards Equation (Richards, 1931), subject to given antecedent soil moisture conditions in the soil profile, the rate of water application at the soil surface, and the conditions at the bottom of the soil profile. In general, the initial soil water potential varies with soil depth, z. The initial conditions) at time t equal 0 and can be expressed as follows

$$h(z, t) = f(z); \quad t = 0 \tag{3.1}$$

where the profile of matric potential head (h) varies with depth (z).

The boundary condition at the soil surface depends upon the rate of water application. For a rainfall event with intensities less than or equal to the saturated hydraulic conductivity of the soil profile, all the rain infiltrates into the soil without generating any runoff. For higher rainfall intensities, all the rain infiltrates into the soil during early stages until the soil surface becomes saturated. After this point, the infiltration is less than the rain intensity and runoff begins. These conditions may be expressed as:

$$-K(h)\frac{\partial h}{\partial z} + 1 = R; \quad \theta(0, t) \leq \theta_s, \; t \leq t_p \tag{3.2}$$

$$h = h_o; \quad \Theta(0, t) = \Theta_s, \; t > t_p, \tag{3.3}$$

where R is the rainfall intensity, h_o is a small positive ponding depth on the soil surface, and t_p is the ponding time. These conditions also accommodate time varying rainfall intensities, as well as when

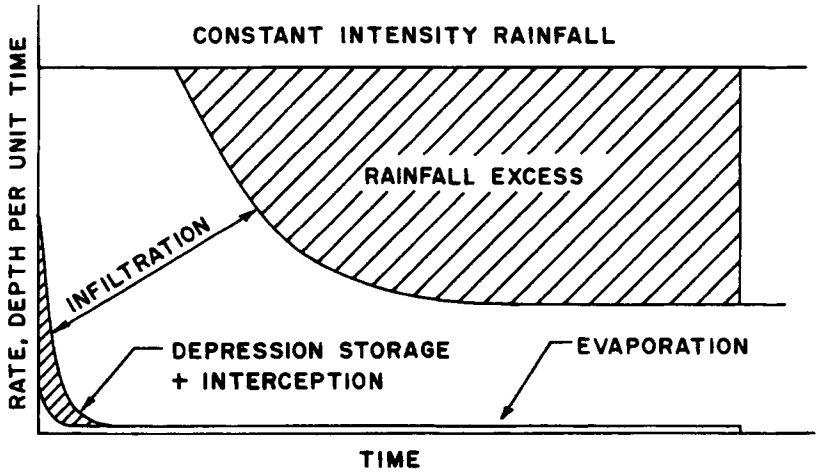


Figure 3.1.—Schematic of Rainfall Excess Components (Sabol et al., 1992).

rainfall intensity is smaller than the saturated hydraulic conductivity (K_s) of the soil throughout the storm. For a surface ponded-water irrigation, condition Equation 3.3 will apply from time zero on.

The surface boundary conditions equations 3.2 and 3.3 apply at any point in the field during rainfall. In a long sloping field, some infiltration may continue to occur in lower parts of the field even after the rainfall stops, due to continued overland flow from upper parts. During this phase, conditions described by equations 3.2 and 3.3 still apply after rainfall is replaced by the overland flow per unit area at the point of interest. To obtain the overland flow rates, hydrodynamic equations of overland flow need to be solved interactively with infiltration (Woolhiser, et al., 1990).

The lower boundary condition depends upon the depth of the unsaturated profile. For a deep profile, a unit-gradient flux condition is commonly applied at a depth, L, below the infiltration-wetted zone:

$$q(L, t) = K(\Theta, L); \quad t > 0. \tag{3.4}$$

For a shallow profile, a constant pressure head is assumed at the water table depth L:

$$h(L, t) = 0; \quad t > 0. \tag{3.5}$$

The Richards Equation 3.1 subject to the general conditions described in Equations 3.2 to 3.5 in a layered soil profile does not have any known analytical or closed-form solutions for infiltration. However, the solutions can be obtained by using finite-difference or finite-element numerical methods (Rubin and Steinhart, 1963; Mein and Larson, 1973).

For non-layered soils, uniform initial soil moisture distribution, and limited surface boundary conditions, some closed-form solutions are available. The pioneering work of Philip (1957) provided a series solution for vertical infiltration into a semi-infinite homogeneous soil, with a constant initial moisture content Θ_i and a constant matric potential h_0 maintained at the soil surface. Recently, Swartzendruber (1987) presented a solution that holds for both small to intermediate and large times.

III. FACTORS AFFECTING INFILTRATION/RAINFALL EXCESS

Factors which affect infiltration have been divided into the following categories: 1) Soil; 2) Surface; 3) Management; and 4) Natural (Brakensiek and Rawls, 1988). Depending on their importance for the specific application, these categories should be accounted for when applying infiltration models. Published research results are used to illustrate the relative importance of the various factors. Methods for incorporating the effects of the factors into infiltration models will be discussed in the Infiltration/Rainfall Excess Models for Practical Applications section.

A. Soil

Soil factors encompass both soil physical properties including particle size, morphological, and chemical properties and soil water properties including soil water content, water retention characteristic, and hydraulic conductivity. Soil water properties are closely related to soil physical properties.

1. Soil Physical Properties

a. Soil Texture. Soil texture is determined from the size distribution of individual particles in a soil sample. Soil particles smaller than 2 mm are divided into three soil texture groups: sand, silt, and clay. Fig. 3.2 shows the particle size, sieve dimension, the defined size class, and the limits for the basic soil texture classes for the U.S. Department of Agriculture (USDA) (Soil Conservation Service, 1982b). The soil texture groups which have the greatest effect on infiltration (Rawls et al., 1991) are the percentages of sand, silt, clay, fine sand, coarse sand, very coarse sand, and coarse fragments (0.2 cm). Fig. 3.3a and 3.3b illustrate that coarse textured soils (gravelly sandy loam) normally have a higher infiltration rate than fine textured soils (clay loam). Coarse fragments in the soil normally increase the infiltration rate of the soil.

b. Morphological Properties. The morphological properties having the greatest effect on infiltration are bulk density, organic matter, and clay type. These properties are closely related to soil structure and soil surface area. As bulk density increases, soil porosity and infiltration decrease. Soil organic matter (1.74

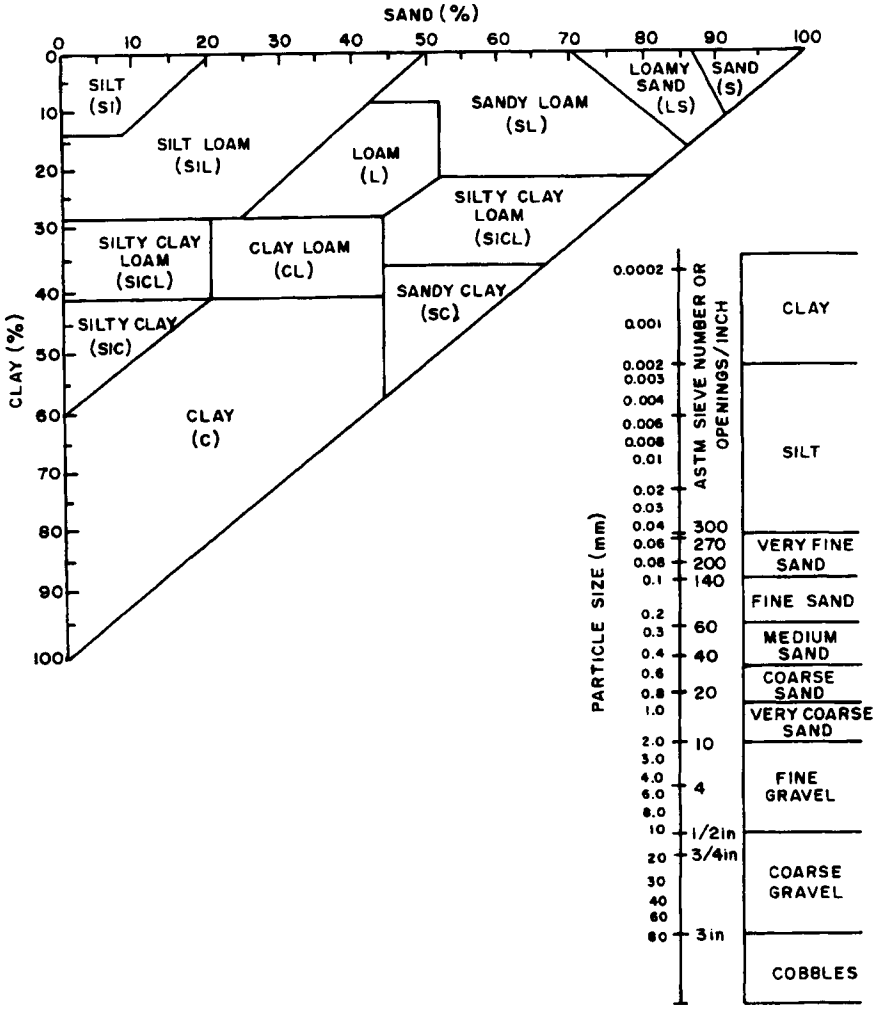


Figure 3.2.—USDA Soil Textural Triangle and Particle Size Limits (SCS, 1986).

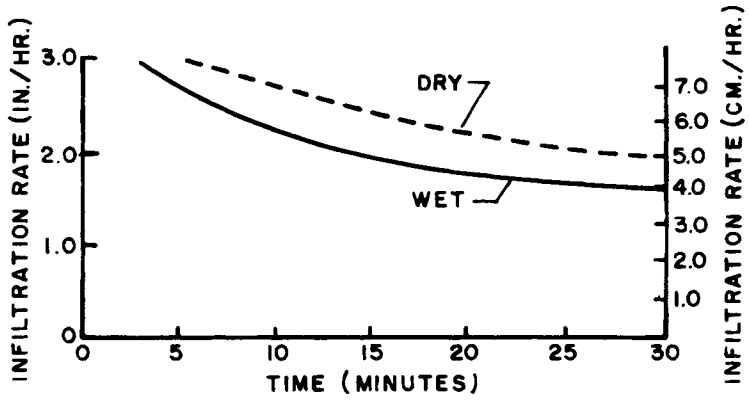


Figure 3.3a.—Infiltration Curves for the Artemisa arbuscula/Poa secunda (low) Community, Coils Creek Watershed, Clay Loam (Blackburn, 1973).

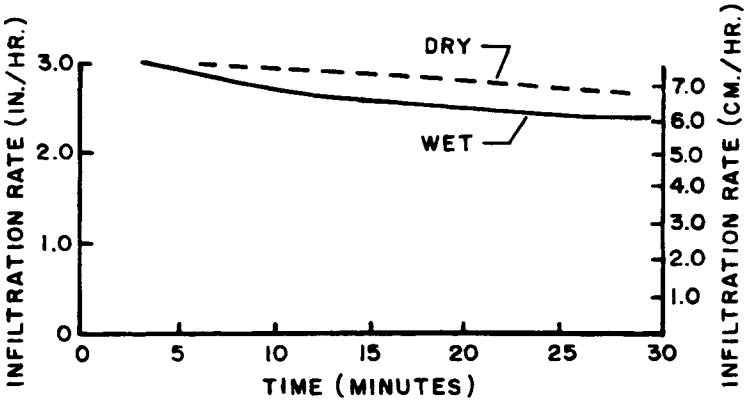


Figure 3.3b.—Infiltration Curves for Symphoricarpos longiflorus/Artemisia tridentata/Agropyron spicatum/Wyethia mollius Community, Coils Creek Watershed, Gravelly Sandy Loam (Blackburn, 1973).

times the percent of organic carbon) is inversely related to bulk density; thus as organic matter increases, infiltration increases.

The clay mineralogy, or clay type, has a significant effect on infiltration for soils containing a large percentage of clay (10%). For example, expandable clays such as montmorillionitic have a significantly lower infiltration rate than nonexpendable clays such as kaolinitic.

Soil layers due to natural profile development, crusts, or compaction limit or modify infiltration rates. Fig. 3.4 illustrates that a soil with a porous layer overlying a less porous layer results in a final infiltration rate that approaches the final rate of the lower layer. Also, the crusted surface, which is synonymous to a less porous layer overlying a more porous layer, produces a significantly lower final infiltration rate.

Macroporosity is natural or man-induced through "channels" connected from the soil surface to some depth in the soil profile. They may be cracks, worm holes, tillage marks, or intentional soil slots. Fig. 3.5 shows that compaction severely reduces the volume of macroscopic porosity and reduces infiltration rates. Natural cracks in a montmorillonitic clay can increase the infiltration rate from 0.05 cm/hr to over 5 cm/hr.

Other morphological properties such as the thickness of the soil horizon and soil structure are derived from soil survey descriptions and a quantitative description of their effects on soil water movement have not been determined.

c. Chemical Properties. Chemical properties of the soil are important because they affect the integrity of the soil aggregates (group of soil particles bound together). Chemical soil properties should be

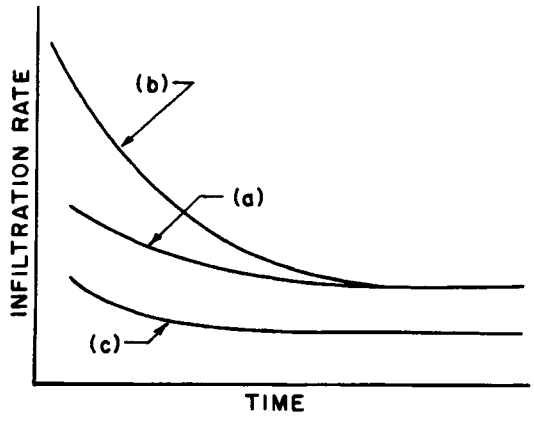


Figure 3.4.—Infiltration Rate for: a) Uniform Soil, b) With a More Porous Upper Layer, and c) Soil Covered With a Surface Crust (Hillel, 1971).

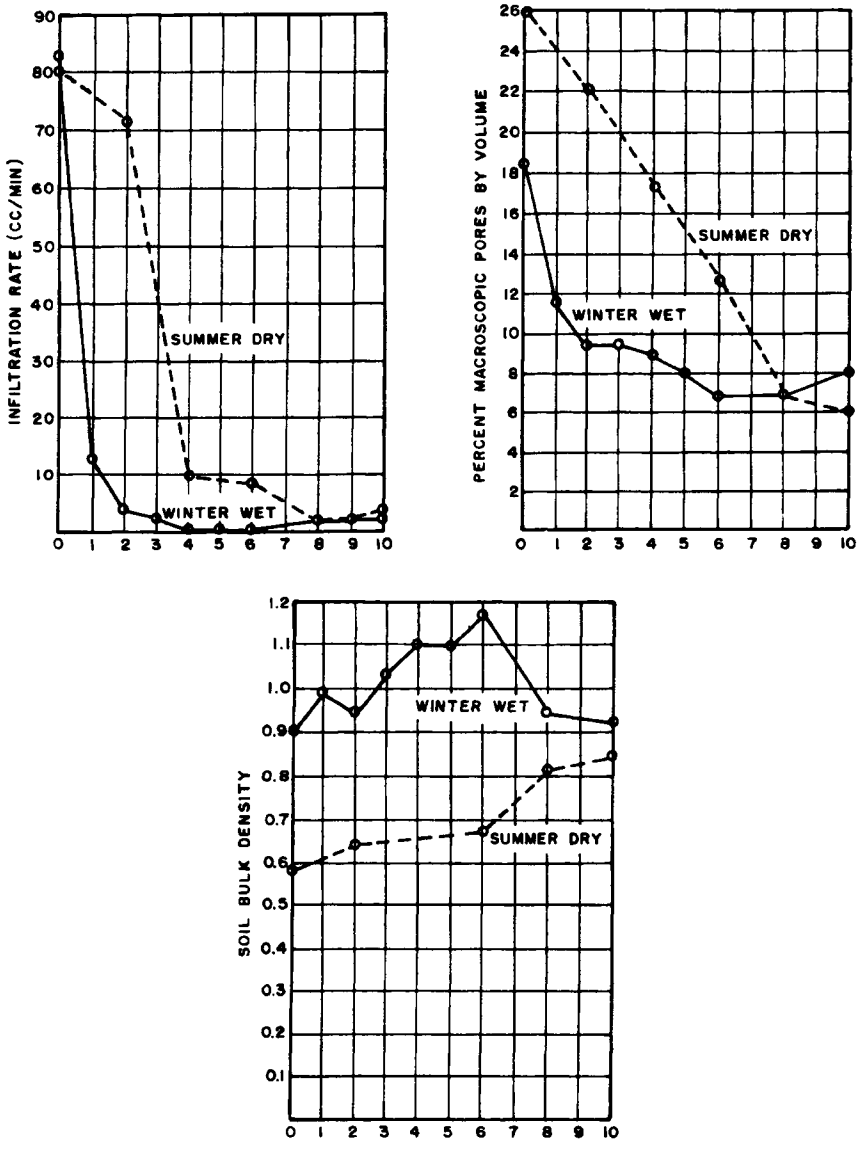


Figure 3.5.—a) Infiltration Rates, b) Volume of Macroscopic Pores, and c) Soil Bulk Density Influenced by Soil; Compaction (Lull, 1959).

considered when they are outside normal ranges (Rawls, et al., 1991). Also, chemistry of the infiltrating water can have an effect on infiltration.

2. *Soil Water Properties*

a. Soil Water Content. Fig. 3.3a and 3.3b illustrate that the higher the water content the smaller the infiltration rate.

b. Water Retention Characteristic. The water retention characteristic of the soil describes the soil's ability to store and release water and is defined as the relationship between the soil water content and the soil suction or matric potential. Fig. 3.6 illustrates the water retention relationship for two contrasting soil textures. Note that the sandy loam soil retains less water than the clay soil.

c. Hydraulic Conductivity. The hydraulic conductivity is the ability of the soil to transmit water and depends upon both the properties of the soil and the fluid (Klute and Dirkson, 1986). Total porosity, pore size distribution, and pore continuity are the important soil characteristics affecting hydraulic conductivity. The hydraulic conductivity at or above the saturation point is referred to as "saturated hydraulic conductivity" and is directly related to infiltration.

The hydraulic conductivity of the sandy loam soil decreases more rapidly with decrease in matrix potential than the hydraulic conductivity of clay soil. Thus, at lower matrix potential (or higher suctions) the hydraulic conductivity of the clay soil is higher. The rate of change of matrix potential in the sandy soil also decreases much more rapidly than that of the clay soil.

Hysteresis is caused by entrapment of air in the soil during wetting, and can cause the hydraulic conductivity to decrease; however, its effect is normally small and for practical applications has mostly been neglected.

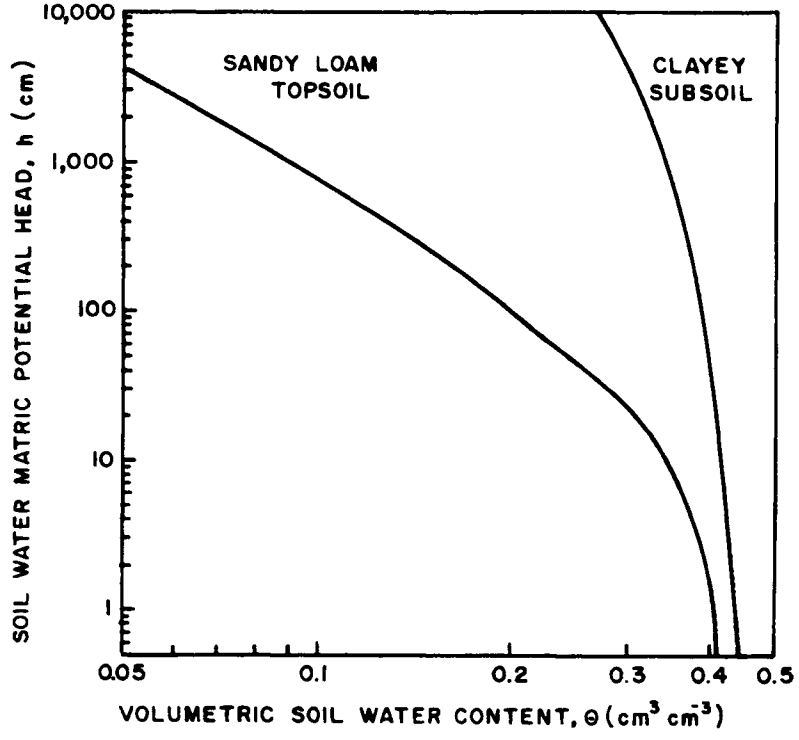


Figure 3.6.—The h(θ) Relationship of Sandy loam and Clayey Horizons of Cecil Soil (Ahuja et al., 1985).

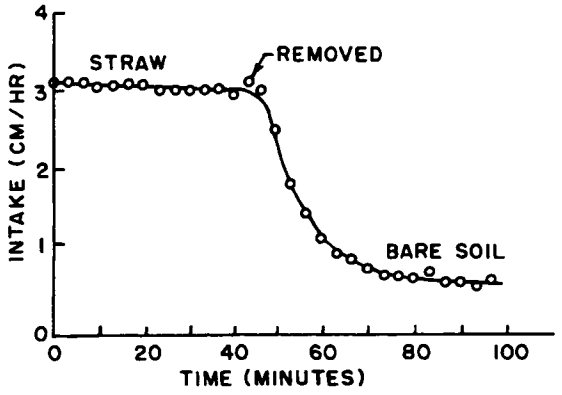


Figure 3.7.—Effect of Straw Cover and Bare Soil on Infiltration Rates (Gifford, 1977).

B. Surface

The surface factors are the factors that affect the movement of water through the air–soil interface. Factors at the soil surface are associated with materials that cover the surface, modify the shape of the surface and remove water by surface storage from its cycle of movement into the soil.

1. Cover Cover factors are materials that protect the soil surface. Lack of cover, or bare soil, is associated with the formation of a surface crust. A surface seal (crust) can develop under the action of raindrops (Summer and Stewart, 1992). The impact of raindrops can break down surface soil structure and cause the movement of soil fines into surface or near surface porosity. Once formed, a crust can impede infiltration. Fig. 3.7 shows the relative effect of surface cover removal on the infiltration rate. Fig. 3.8 indicates a strong relationship between total organic cover and infiltration rates. Fig. 3.9 exhibits the drastic effect of a surface crust on infiltration.

The effect that bare rangeland soil has on infiltration is shown in Fig. 3.10. Note the interaction with soil moisture. Fig. 3.11 indicates the effect of shrub canopy to increase infiltration. Fig. 3.12 points out the effect of surface cover, gravel, bare, litter, or crown cover on total infiltration. Bare and gravel cover reduce infiltration. Figs. 3.13 and 3.14 show that a grass or rock cover decreases the effect of trampling to reduce infiltration. The effect of cover, clipped cover, and bare ground on the final infiltration rate is shown in Fig. 3.15. Surface cover protects the ground surface from crusting and enhances infiltration rates.

2. Configuration Surface configurations can be natural or man-made. Fig. 3.16 is an example of the effect of natural soil mounds under shrubs on rangeland in Nevada. Compared to interspace areas, the dune areas have much higher infiltration rates. The soil of the dune areas is higher in organic matter and has a lower bulk density (higher porosity).

Man-made configurations result from various kinds of tillage. In tillage, this is referred to as surface roughness and is measured as random roughness (Zobeck and Onstad, 1987). Table 3.1 (Freebairn, 1989) exhibits the random roughness change due to several kinds of tillage on Minnesota silt loam soil. Note that tillage also reduces bulk density. Fig. 3.17 shows the tendency for an increasing random roughness to be associated with higher infiltration rates. The effect of roughness due to tillage is enhanced when protected from rainfall impact. Fig. 3.18 shows the surface roughness effect due to clods on infiltration and how this influence declines with exposure to rainfall. The dashed line, referring to surface cover with a furnace filter, is evidence of the efficiency of cover to sustain infiltration rates.

3. Storages Surface storage results from interception and/or depressions. Interception is caused by vegetation or other types of surface cover. Tables and equations for agricultural areas are found in Horton (1919) and Onstad (1984). Both interception and depressions reduce rainfall excess and may be very significant for small storm events.

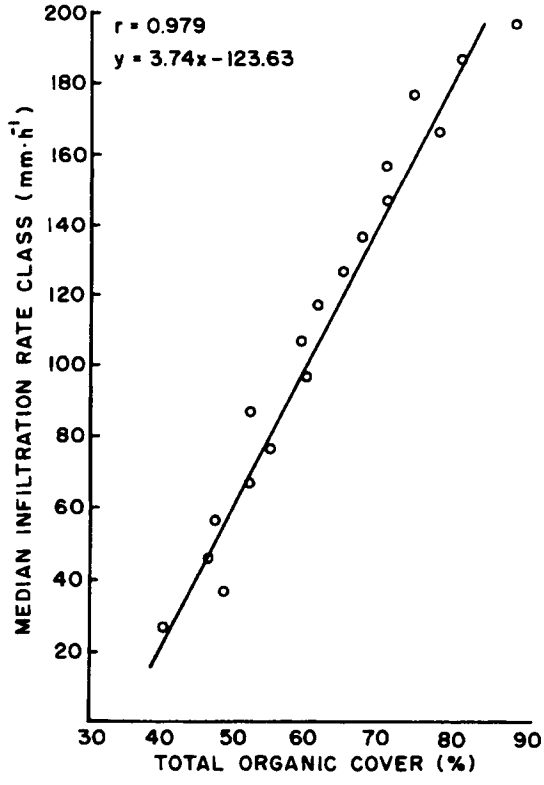


Figure 3.8.—Relationship of Median Infiltration Rate Classes with Total Organic Cover (%), Edwards Plateau, TX (Thurow, 1985).

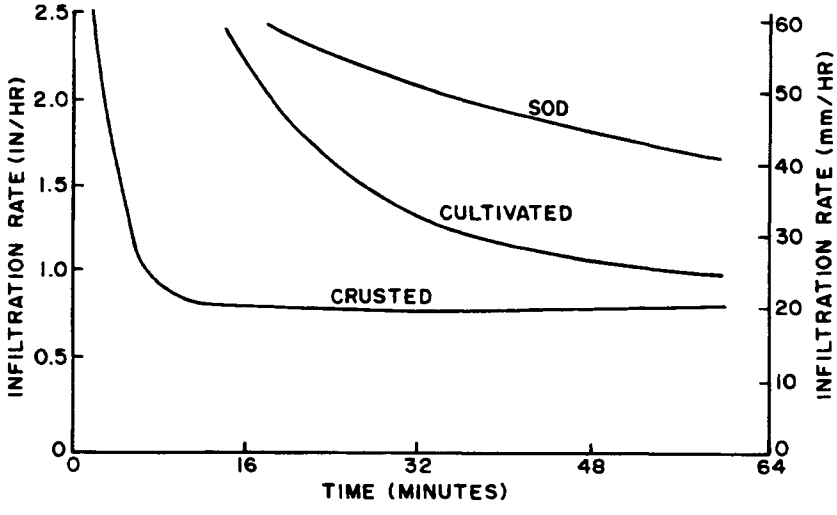


Figure 3.9.—Effect of Surface Sealing and Crusting on Infiltration Rate for a Zanesville Silt Loam (Skaggs and Khaleel, 1982).

C. Management

Management, as a factor affecting infiltration, modifies the soil properties or the soil surface condition. The effects of these factors have already been discussed in previous sections. In this section, the effect of management practices on infiltration will be presented.

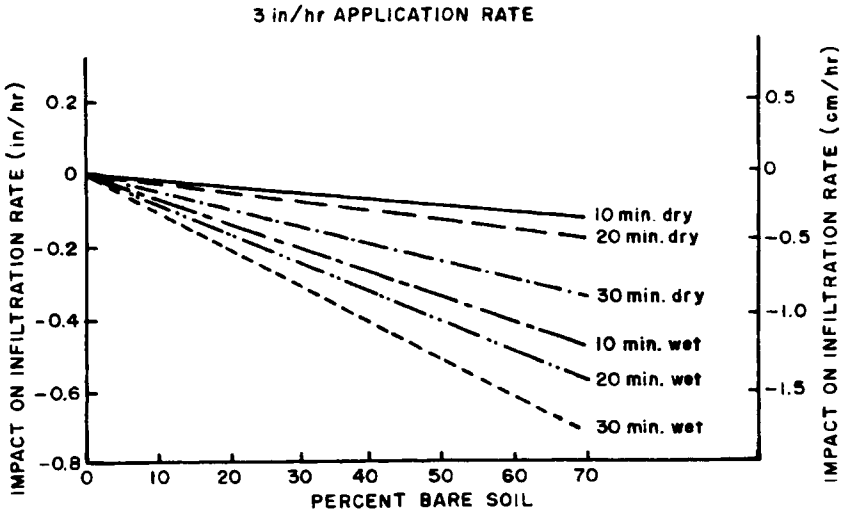


Figure 3.10.—Impact of Percent Bare Soil on Infiltration Rates After Various Time Intervals (Gifford, 1977).

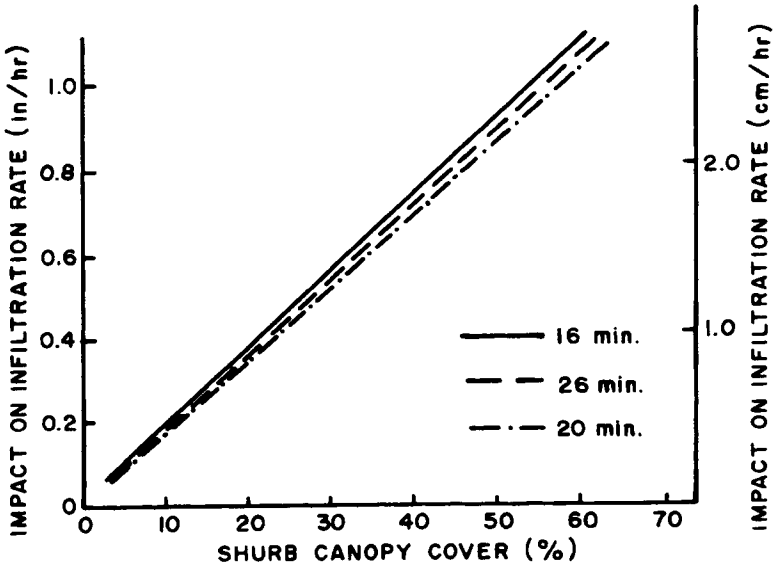


Figure 3.11.—Impact on Shrub Canopy on Infiltration Rates After Various Time Intervals (Gifford, 1977).

1. Agriculture Agricultural management systems involve different types of tillage, vegetation, and surface cover. Fig. 3.19 illustrates tillage practices (moldboard plow, chisel plow, and no till) influence on infiltration. Brakensiek and Rawls (1988) reported that moldboard plowing would increase soil porosity from 10 to 20% depending on soil texture and hence, increase infiltration rates over non-tilled soils. Rawls (1983) reported that increasing the organic matter of the soil lowers the bulk density, increases porosity and hence, increases the infiltration. Fig. 3.20 gives the steady-state infiltration rate of bare ground, soybean vegetation, and residue cover interactions for planting, midseason, and harvest. The bare soil steady-state rate decreases between planting and midseason and remains stable primarily as a result of crusting. Residue maintains a high steady-state rate until harvest, while the canopy and canopy residue combination increase the steady-state rate.

Table 3.2 (North Central Regional Committee 40, 1979) presents a grouping of major Midwest agricultural soils under tillage or grass management. There is a tendency for the same soil under grass

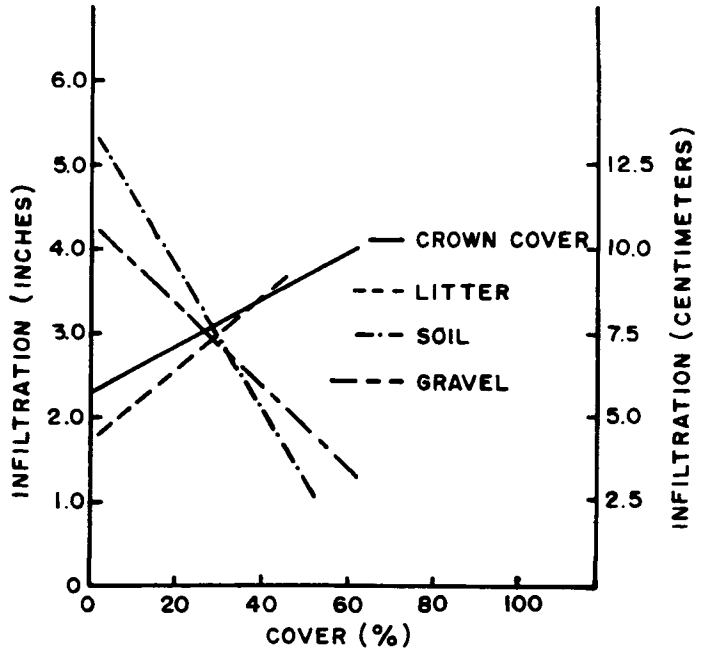


Figure 3.12.—Relationship of Various Covers to Infiltration Rates (Gifford, 1977).

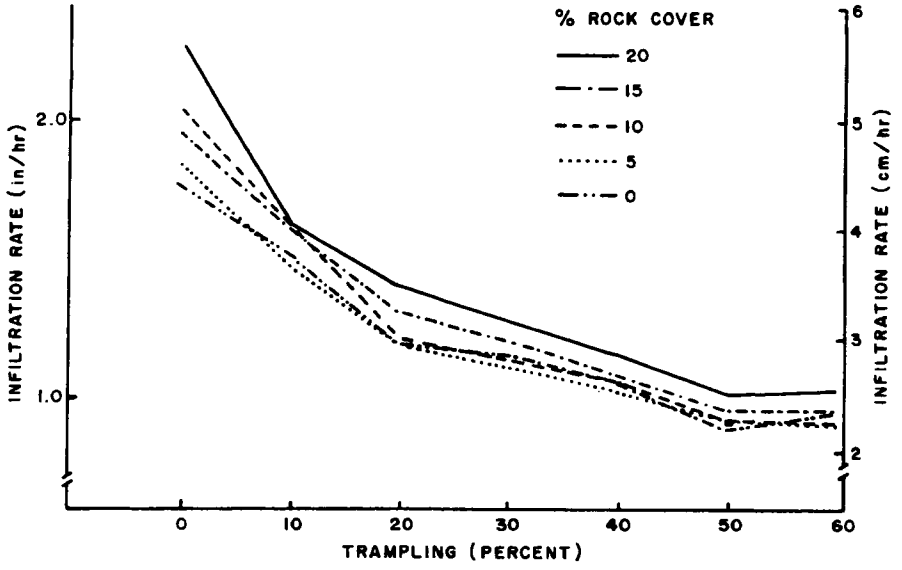


Figure 3.13.—Infiltration Rate at 30 Minutes as a Function of Rock Cover and Trampling (Dadkah and Gifford, 1980).

management to have higher infiltration rates than when under tilled management. Fig. 3.19 clearly depicts the influence of tillage management on infiltration rates. Fig. 3.21 indicates that the influence of tillage management decreases as the surface is exposed to rainfall. A surface cover reduces the degradation of the tillage affect. Rawls et al. (1983) developed a diagram from tillage data for estimating the effect of plowing to increase soil porosity, hence, increasing infiltration rates.

Additions to the soil that increase organic matter would increase infiltration. As shown in Fig. 3.11, the soil area under shrubs where litter accumulates and soil organic matter increased supported higher

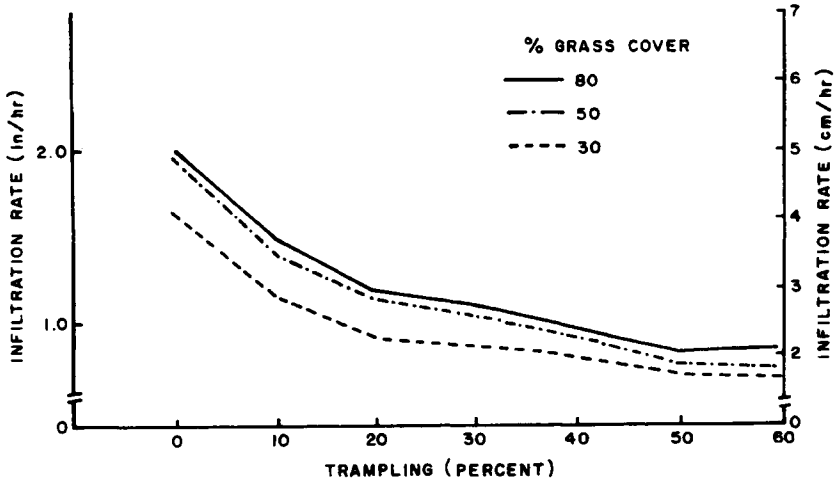


Figure 3.14.—Infiltration Rate at 30 Minutes as a Function of Grass Cover and Trampling (Dadkhah and Gifford, 1980).

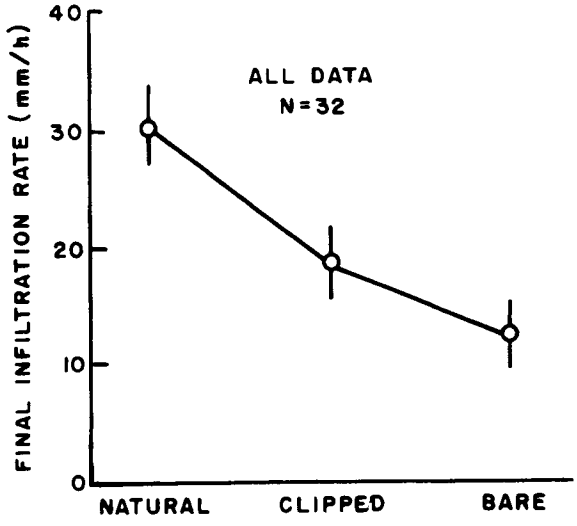


Figure 3.15.—Final Infiltration Rates Shown as a Mean From Five Sites in Arizona and Nevada for Very Wet Runs with Natural, Clipped, and Bare Cover (Lane et al., 1987).

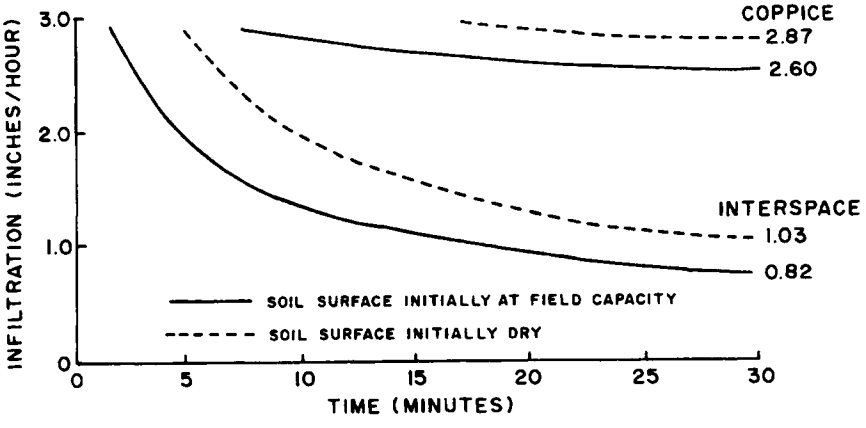


Figure 3.16.—Infiltration Curves for the Artemisia tridentata Community, Dickwater Watershed, Coppice Dune and Interspace (Blackburn, 1973).

TABLE 3.1. Influence of Tillage on Random Roughness (Freebairn, 1989).

Port Byron silt loam-Lawler farm. Rochester, MN					
Tillage meth	Plot	RR before tillage mm	RR after tillage mm	Bulk den before gm/cm³	Bulk den after gm/cm³
Chisel	E	9	17	1.16	1.07
plow	F	9	18	1.16	1.07
	G	9	16	1.16	1.07
	H	9	21	1.16	1.07
Moldboard	I	9	22	1.16	0.97
plow	J	9	19	1.16	0.97
	K	9	17	1.16	0.97
	L	9	25	1.16	0.97
Fall chisel	AA	7	24	1.16	1.15
	BB	7	18	1.16	1.15
Fall plow	EE	7	42	1.16	0.99
	FF	7	27	1.16	0.99

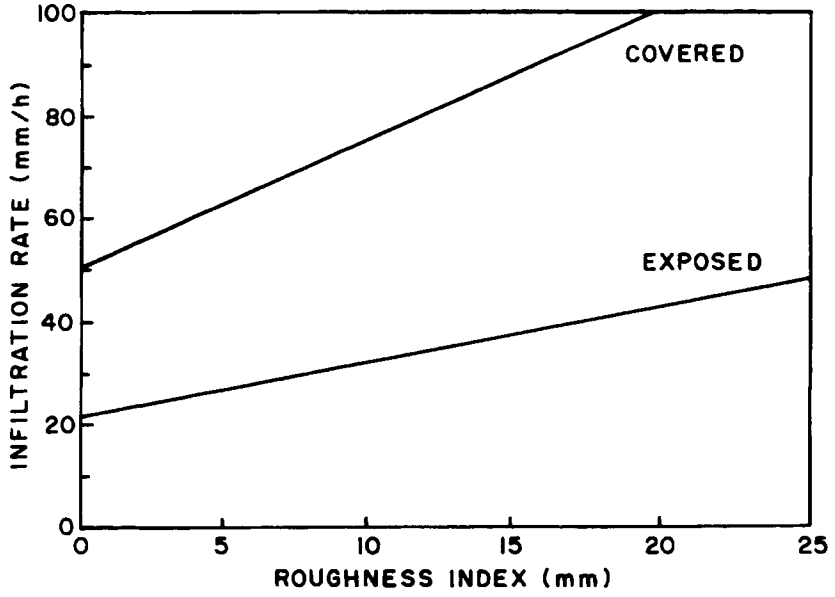


Figure 3.17.—Infiltration Rate vs. Roughness Index for Covered and Exposed Agricultural Plots (Freebairn, 1989).

infiltration rates. Rawls (1983) reports organic matter to be a major factor that lowers the bulk density (increased porosity); hence, increasing organic matter increases the hydraulic conductivity.

2. ***Irrigation*** Many of the factors in previous sections apply to infiltration under irrigation practices. Sprinkler systems are similar to rainfall infiltration processes. Border irrigation is similar to ponded infiltration processes. Furrow irrigation factors have been investigated by Trout and Kemper (1983) and by Kemper et al. (1988). Sub-irrigation systems involve factors common to soil moisture movement.

3. ***Rangeland*** The type of vegetation on rangelands is an important factor determining infiltration rates. Fig. 3.22 reveals a major difference of mean infiltration rates for three vegetation types. In Fig. 3.23, a significant effect on total infiltration is seen between successional stages of rangeland vegetation. Fig.

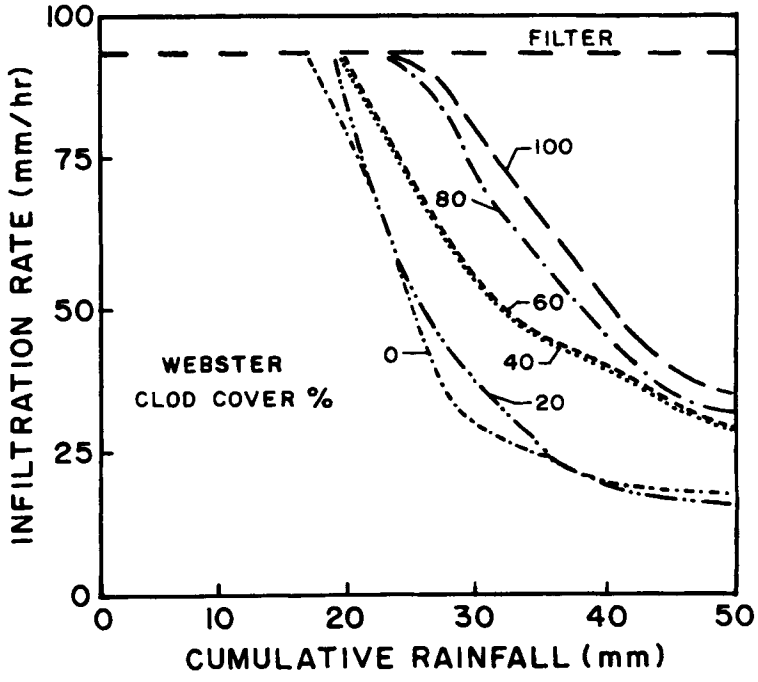


Figure 3.18.—Infiltration Rates Webster Clay Loam with the Surface Covered with Various Percentages of 25-50 mm Clods or Furnace Filter (Freebairn, 1989).

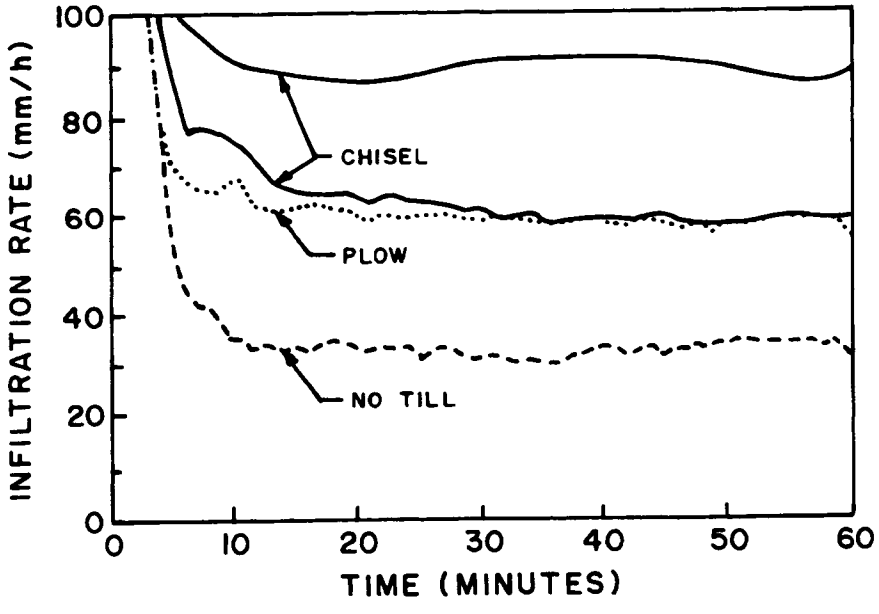


Figure 3.19.—Infiltration Rates for Port Byron Silt Loam, Two Months After Chisel and Moldboard Plow Tillage and Four Months After Planting Under No-till (Freebairn, 1989).

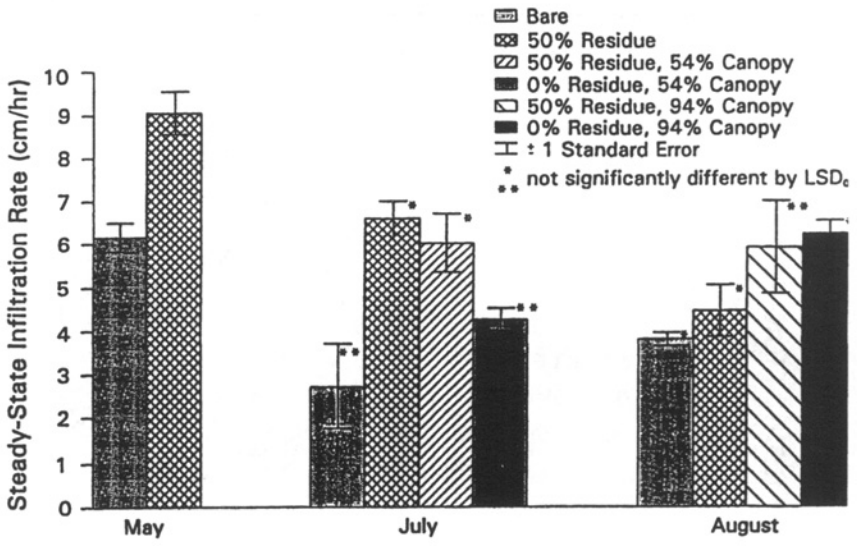


Figure 3.20.—Seasonal Effects of Agricultural Practices on Steady-State Infiltration Rates (Rawls et al., 1993a).

3.24 shows an improvement of infiltration rates going from bare ground to grass to shrub vegetation types.

Grazing practices also influence infiltration. Fig. 3.25 provides evidence that grazing reduces the final infiltration rate. Heavy grazing is associated with a greater reduction of the final infiltration rate. From studies in New Mexico, presented in Fig. 3.26, a difference is reported between several grazing treatments. Fig. 3.27 clarifies the effect of stocking density on infiltration rates. The effect of grazing systems on infiltration is shown in Fig. 3.28. The three grazing systems are MCG-moderate continuous, HCG-heavy continuous, and SDG-short duration grazing. Effects of rangeland improvements on runoff are summarized in Rangeland Hydrology (Branson et al., 1981). Fig. 3.29 shows an effect of burning on infiltration rates.

D. Natural

Natural factors include natural processes such as precipitation, freezing, change of seasons, temperature, and moisture which vary with time and space and interact with other factors in their effect on infiltration. The temporal and spatial variability effects will be discussed.

1. Temporal The effect of cumulative antecedent rainfall on exposed and 50% residue-covered agricultural soil is shown in Fig. 3.20, indicating a decrease in the steady-state infiltration rate with continued exposure to the action of rainfall. For bare soil, it seems that a stable steady-state infiltration rate is achieved between planting and midseason, indicating that a stable crust is achieved early in the growing season and maintained thereafter. Fig. 3.20 demonstrates that the steady-state infiltration rate increases as canopy cover increases over the growing season. Also, Fig. 3.20 indicates that canopy cover and residue cover do not cause additive increases in the steady-state infiltration rate.

Increases in rainfall intensity expand surface disturbance caused by the rain drops and the buildup of a ponding head. This usually increases the bare soil infiltration. For bare soil with canopy cover, this intensity effect is dissipated by the growing crop canopy.

Soil temperature influences infiltration through its effect on the viscosity of water. Lee (1983) found that freezing the soil with a high moisture content decreases infiltration to almost zero, while freezing the soil at a low moisture content increases infiltration by twice its normal rate.

TABLE 3.2. Alphabetical Listing of Some Major Soils of the North Central Region of the United States and Alaska and Their Mean Equilibrium Infiltration Rates, As Measured with a Sprinkling Infiltrometer (North Central Regional Committee 40, 1979).

Soil	Rate (inches per hour)		Test location
	Tilled surface	Grass surface	
Bearden silty clay loam	.2	.9	N. Dakota
Blount silty clay loam	.3	.3	Ohio
Bodenburg silt loam	.4	.4	Alaska
Canfield silt loam	.4	1.1	Ohio
Cincinnati silt loam	1.1	1.1	Indiana
Cisne silt loam	.6	.6	Illinois
Clermont silt loam	.8	.8	Indiana
Dudley clay loam	.4	1.0	S. Dakota
Elliott silt loam	1.5	1.5	Illinois
Emmet loamy sand	1.5	1.5	Michigan
Fayette silt loam	.2	.6	Wisconsin
Flanagan silt loam	1.9	3.6	Illinois
Grundy silt loam	1.1	1.1	Iowa
Holdrege silt loam	1.2	1.2	Nebraska
Hoytville clay	.3	1.0	Ohio
Ida silt loam	.7	1.5	Iowa
Keith very fine sandy loam	.7	1.5	Nebraska
Kenyon silt loam	1.3	1.3	Minnesota
Knik silt loam	.4	.4	Alaska
Kranzburg silty clay loam	.8	1.0	S. Dakota
Miami sandy loam	1.2	1.2	Michigan
Minto silt loam	.4	.4	Alaska
Moody silt loam	.4	1.4	Iowa
Morton loam	.4	1.0	N. Dakota
Plainfield loamy sand	3.7	4.4	Wisconsin
Port Byron silt loam	.5	2.0	Minnesota
Russell silt loam	.5	.5	Indiana
Sharpsburg silty clay loam	1.3	1.3	Nebraska
Sims loam	1.1	2.5	Michigan
Sinai silty clay loam	.6	1.1	S. Dakota
Webster clay loam	1.0	.6	Minnesota
Withee silt loam	.2	1.1	Wisconsin

Fig. 3.30 shows a seasonal influence of native rangeland infiltration rates. Fig. 3.20 indicates a seasonal effect for agricultural lands on the final infiltration rates.

2. ***Spatial*** The effect of spatial variability on infiltration is a measure of the difference between "point" infiltration and apparent infiltration rates associated with composite areas or watersheds. The variability of soil, surface, and management factors over an area is a result of the same practices or factors that have been discussed in previous sections. Smith (1983) states "there is no set of parameters which, when used with point infiltration models will produce the same response as the net (ensemble) from the area."

Smith and Hebbert (1979) studied the effect of a linear variation of hydraulic conductivity along a plane "watershed" on runoff for two rates of uniform rainfall. Their results clearly indicate the effect of different spatial conductivity patterns on runoff hydrographs, i.e., rainfall excess. It is also apparent that a strong interrelationship is at work between the spatial infiltration patterns and rates of rainfall. A random infiltration pattern tends to occur on the plane which produces a runoff hydrograph similar to a uniform pattern of infiltration for at least higher rainfall rates. However, for lower rainfall rates, the impact of random infiltration variability on runoff (rainfall excess) is still a major factor.

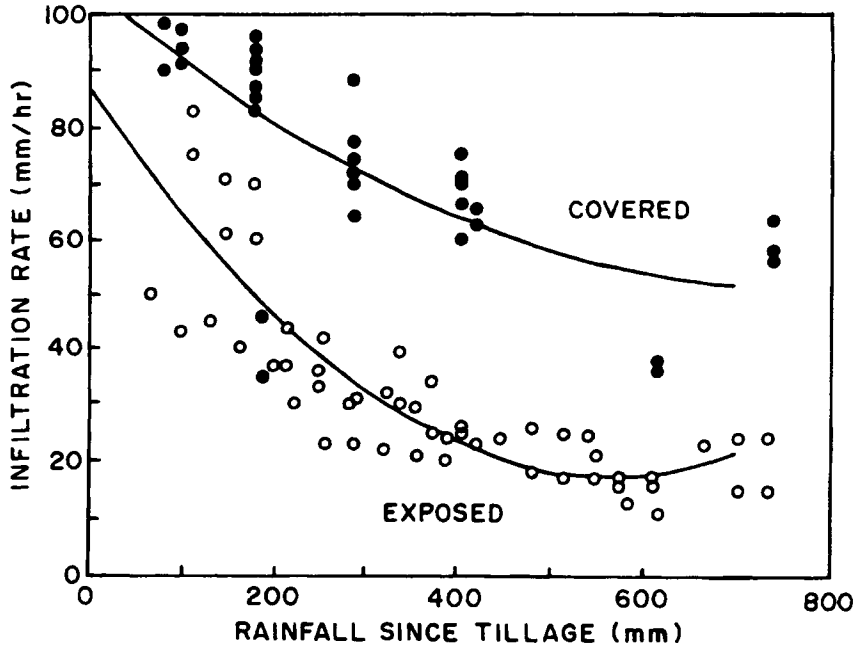


Figure 3.21.—Infiltration Rates vs. Antecedent Rainfall Since Tillage for Covered and Exposed Plots, Webster Clay Loam (Freebairn, 1989).

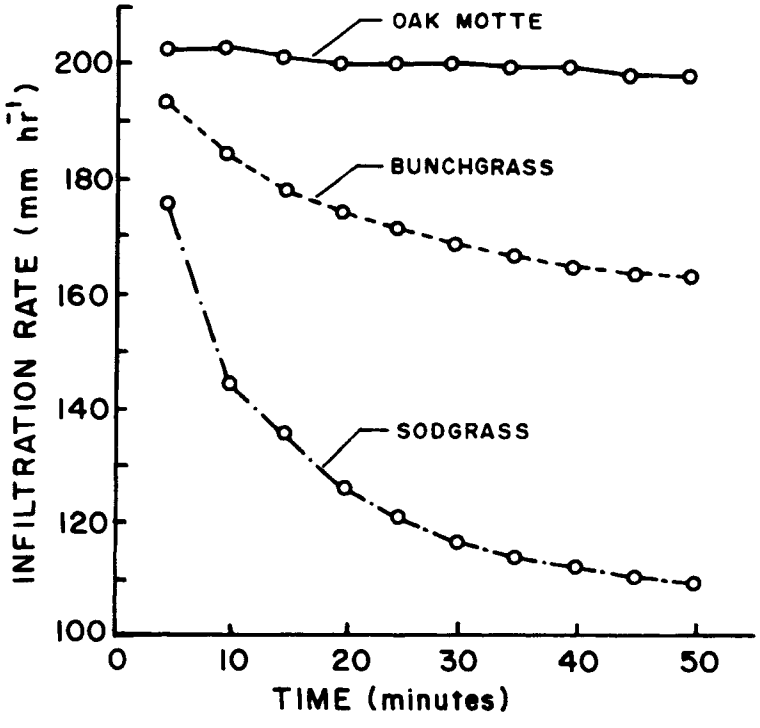


Figure 3.22.—Mean Infiltration Rates for Three Vegetation Types, Edwards Plateau, TX (Thurow et al., 1986).

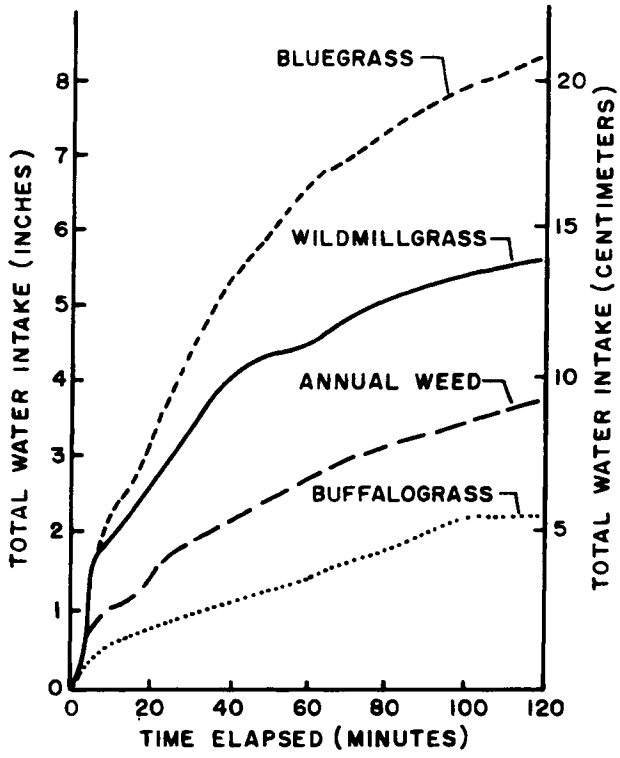


Figure 3.23.—Infiltration Amounts for Four Successional Stages of Rangeland Vegetation (Gifford, 1977).

III. INFILTRATION/RAINFALL EXCESS MODELS FOR PRACTICAL APPLICATIONS

Infiltration models for field applications usually employ simplified concepts which predict the infiltration rate or cumulative infiltration volume. This assumes that surface ponding begins when the surface application rate exceeds the soil surface infiltration rate. The rainfall excess models that lump all losses (infiltration, depression storage, interception) are strictly empirical models.

A. Rainfall Excess Models

Rainfall excess is the part of rainfall that is not lost to infiltration, depression storage, and interception. While a number of models have been proposed for estimating rainfall excess, the most commonly-used models are the index models and the SCS curve number model.

Index models are relatively simple methods that can be useful when performing gaged analysis (e.g., when rainfall and runoff data are available) or when a simple method is commensurate with the data available for estimating loss values. The most commonly used index models are 1) phi index, 2) initial and constant loss rate and 3) constant proportion loss rate (Pilgrim and Cordery, 1992).

1. Phi Index The phi index, ϕ, is probably the most widely used index model. The loss rate defined by the phi index is shown in Fig. 3.31 and is described mathematically as

$$f(t), = I(t), \quad \text{for } I(t) < \phi, \tag{3.6}$$

where f(t) is the loss rate, I(t) is the rainfall intensity, t is time, and ϕ is a constant called the ϕ index. The phi index can be estimated from storm data by separating baseflow from the total runoff and then determining the ϕ which causes the rainfall excess to equal the total runoff. The advantage of this method

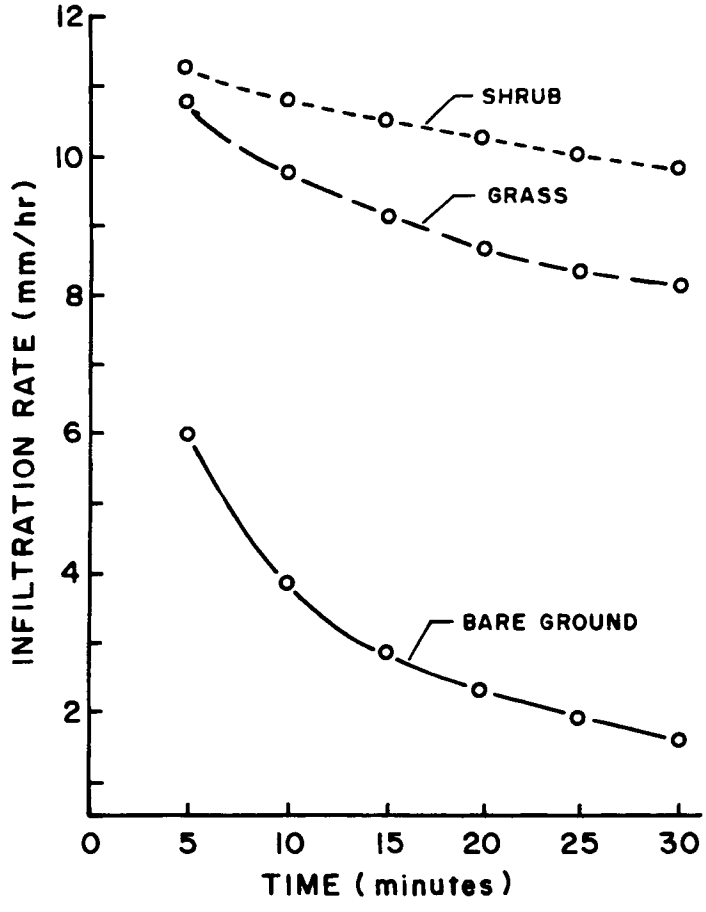


Figure 3.24.—Mean Infiltration Rates for Shrub (GRBI), Grass (CHRO), and Bare Ground (BAGR) (Mbakaya, 1985).

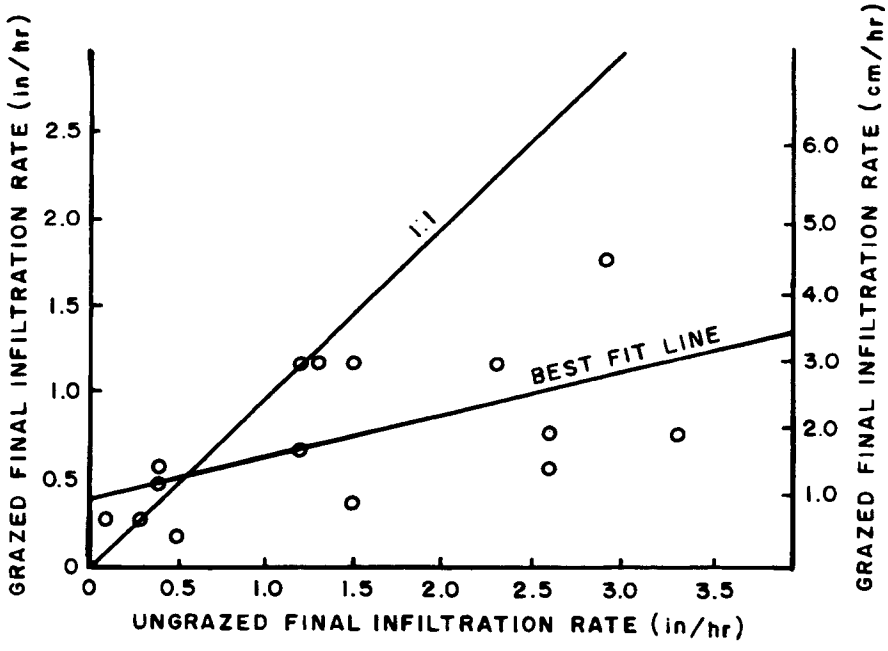


Figure 3.25.—Relationships Between Final Infiltration Rates on Heavily Grazed Areas (Gifford and Hawkins, 1978).

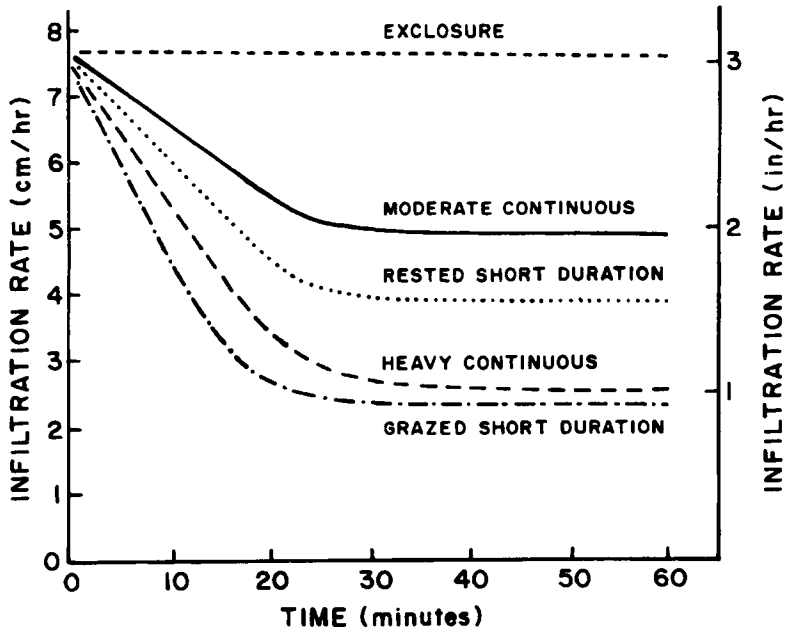


Figure 3.26.—Mean Infiltration Rates for Various Grazing Treatments at Fort Stanton, NM (Weltz and Wood, 1986).

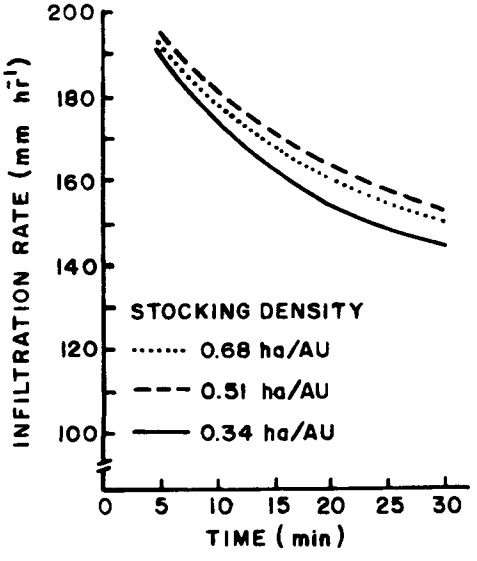


Figure 3.27.—Mean Infiltration Rates for Pastures Grazed at Three Stocking Densities (Warren et al., 1986).

is that it requires only a single parameter. The disadvantages are that it requires rainfall runoff records and the ϕ is dependent on the watershed and storm conditions from which it was determined.

2. Initial and Constant Loss Rate The loss rate function associated with this procedure is shown in Fig. 3.32 and is described mathematically as

$$\begin{aligned} f(t) &= I(t) \quad \text{for } P(t) < IA \\ f(t) &= I(t) - C \quad \text{for } I(t) > C,\ P(t) \geq IA \\ f(t) &= I(t) \quad \text{for } I(t) \leq C, \end{aligned} \tag{3.7}$$

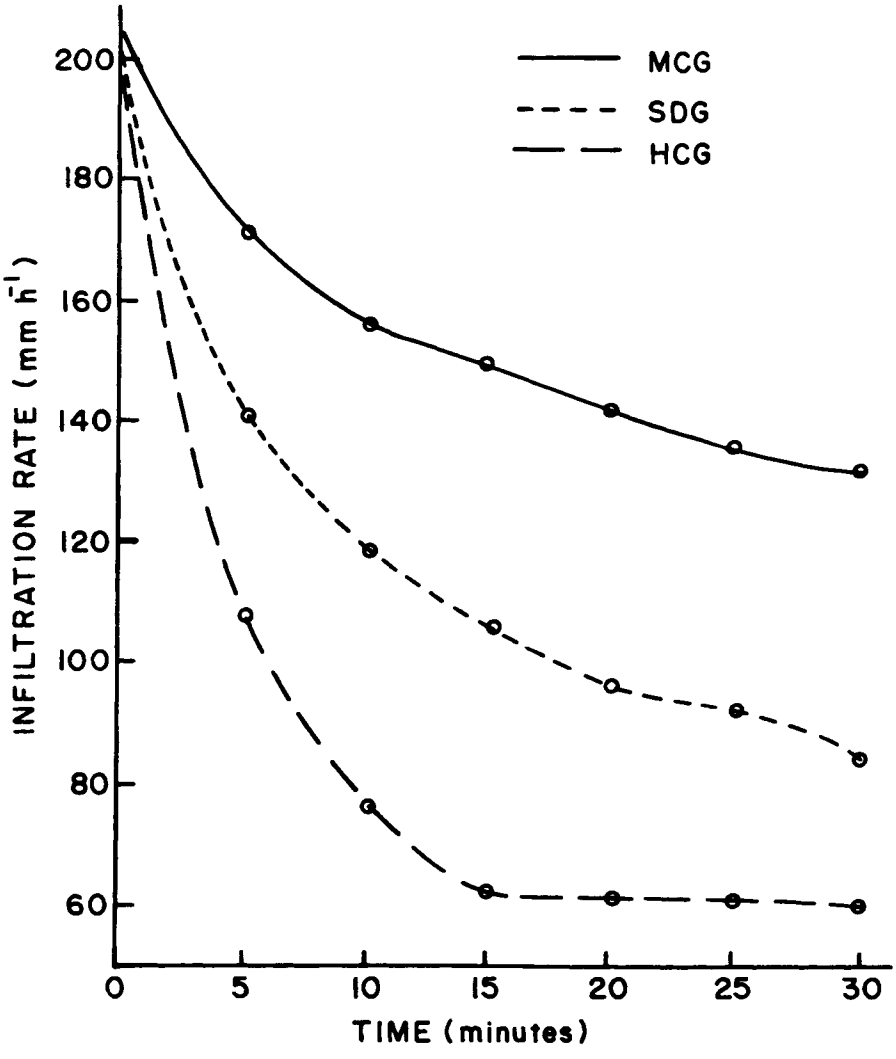


Figure 3.28.—Infiltration Rates for Three Grazing Systems on the Edwards Plateau: MCG-Moderate Continuous, HCG-Heavy Continuous, and SDG-Short Duration (Thurow, 1985).

where P(t) is the cumulative rainfall volume at time t from the beginning of rainfall, I(t) is the rainfall intensity, IA is the initial loss, and C is the constant loss rate. This method is a crude approximation to a typical infiltration curve that decays from some initial high rate to a final constant infiltration rate. The initial loss might be considered to represent the total loss due to surface factors and volume infiltrated prior to attaining the soils long-term infiltration rate. This method is similar to the phi index and has similar advantages and disadvantages. Table 3.3 presents some currently used loss rates (Sabol et al., 1992).

3. Constant Proportion Loss Rate The loss rate function associated with this procedure is shown in Fig. 3.33 and is described mathematically as

$$f(t) = CP * I(t), \tag{3.8}$$

where f(t) is the loss rate, I(t) is rainfall intensity, and CP is a constant ranging from 0 to 1. The advantages and disadvantages are the same as the phi index; however, it does not realistically represent the infiltration process. Calculation of CP requires rainfall and runoff records.

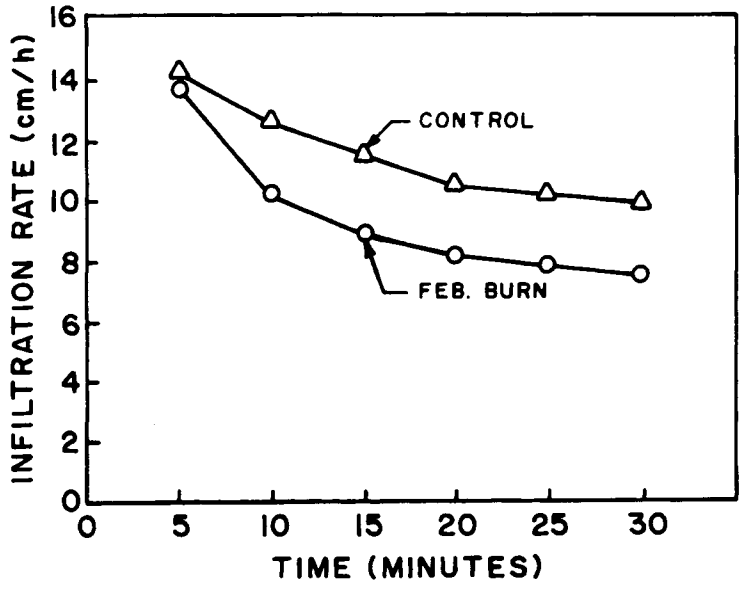


Figure 3.29.—Mean Infiltration Rates of February Burn and Control Areas, Kinoko, Kenya (Cheruiyot, 1984).

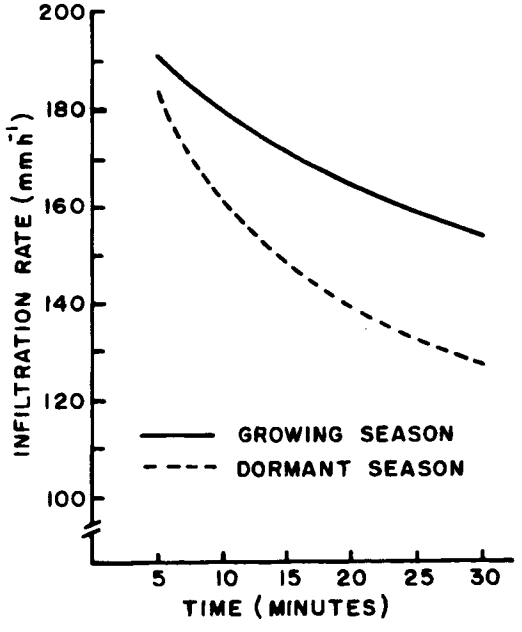


Figure 3.30.—Mean Infiltration Rates During the Growing and Dormant Seasons Near Sonora, TX (Warren et al., 1986).

4. SCS Runoff Curve Number Model The SCS Runoff Curve Number (CN) method is described in detail in Chapter 4 of the Soil Conservation Service National Engineering Handbook (1972). The SCS runoff equation is

$$Q = \frac{(P - I_a)^2}{(P - I_a) + S}, \tag{3.9}$$

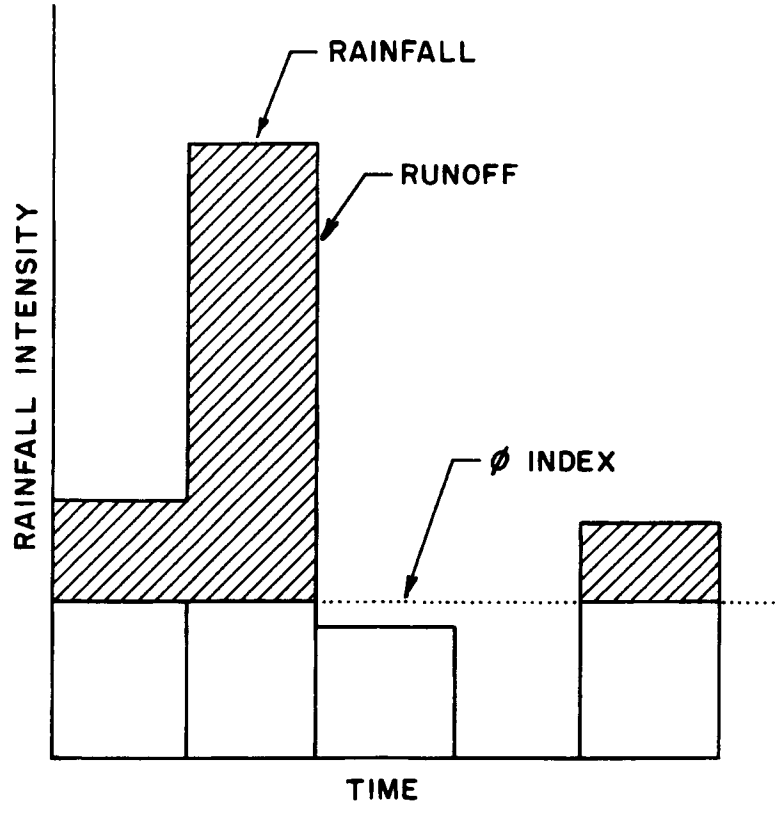


Figure 3.31.—ϕ Index Loss Rate.

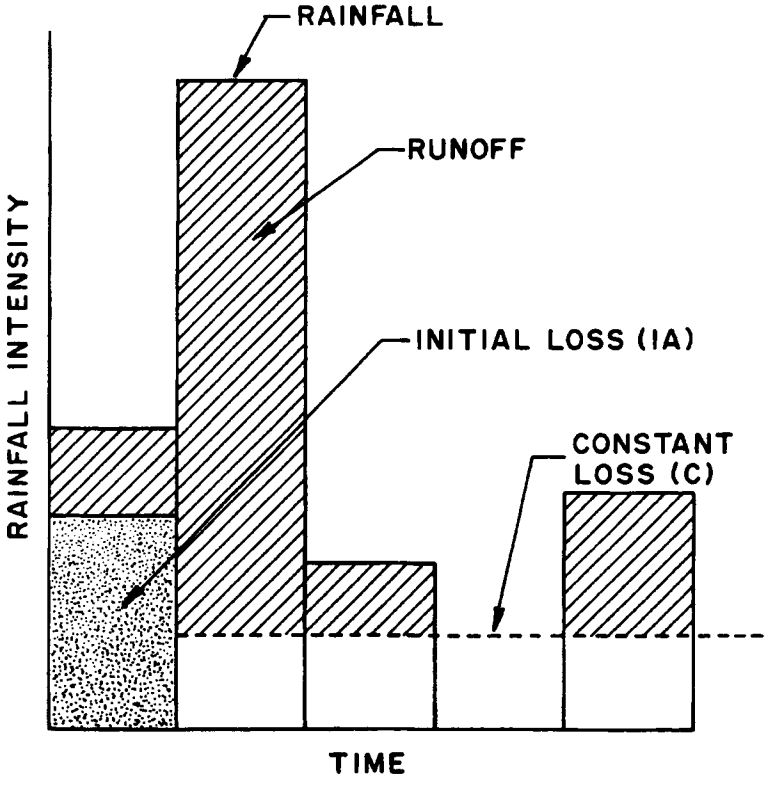


Figure 3.32.—Initial and Constant Loss Rate.

TABLE 3.3. Constant Loss Rates (Sabol et al., 1992).

Hydrologic soil group[1]	Uniform Loss Rate, Inches/hour	
	Musgrave (1955)	U.S. Bureau of Reclamation (1987)
A	0.30–0.45	0.30–0.50
B	0.15–0.30	0.15–0.30
C	0.05–0.15	0.05–0.15
D	0.00–0.05	.0–0.05

1. U.S. Soil Conservation Service

where Q is the runoff (in), P is the rainfall (in), S is the potential maximum retention after runoff begins (in), and I_a is the initial abstraction (in).

Initial abstraction is all losses before runoff begins. It includes water retained in surface depressions and water intercepted by vegetation, evaporation, and infiltration. I_a is highly variable but according to data from many small agricultural watersheds, I_a was approximated by the following empirical equation:

$$I_a = 0.2S. \tag{3.10}$$

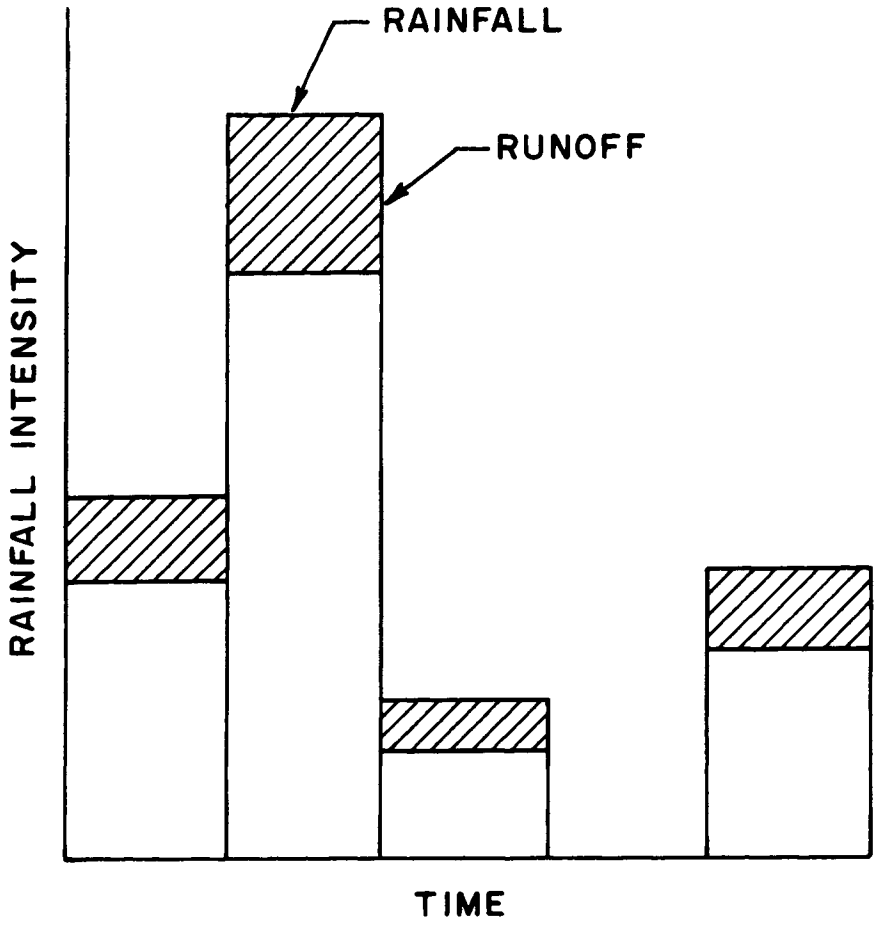


Figure 3.33.—Constant Proportion Loss Ratio.

By eliminating I_a as an independent parameter, this approximation allows the use of a combination of S and P to produce a unique runoff amount. Substituting Equation 3.10 into Equation 3.9 gives

$$Q = \frac{(P - 0.2S)^2}{(P + 0.8S)} \tag{3.11}$$

where the parameter S is related to the soil and cover conditions of the watershed through the curve number, CN. A graphical solution of Equation 3.11 is given in Fig. 3.34. CN has a range of 40 to 100, and S is related to CN by

$$S = \frac{1000}{CN} - 10. \tag{3.12}$$

Equation 3.12 calculates S in units of inches. The major factors that determine CN are the hydrologic soil group, cover type, treatment, hydrologic condition, and antecedent runoff condition. The values of CN's in Table 3.4 (a to d) represent average antecedent runoff conditions for urban, cultivated agricultural, other agricultural, and arid and semiarid rangeland uses (Soil Conservation Service, 1986). The following sections explain how to determine factors affecting the CN.

a. Hydrologic Soil Groups The SCS has classified all soils into four hydrologic soil groups (A, B, C, and D) according to their infiltration rate, which is obtained for bare soil after prolonged wetting. The four groups are defined as follows.

Group A soils have low runoff potential and high infiltration rates even when thoroughly wetted. They consist chiefly of deep, well- to excessively-drained sands or gravels. The USDA soil textures normally included in this group are sand, loamy sand, and sandy loam. These soils have a transmission rate greater than 0.3 in/hr.

Group B soils have moderate infiltration rates when thoroughly wetted and consist chiefly of moderately deep to deep, moderately well- to well-drained soils with moderately fine to moderately coarse

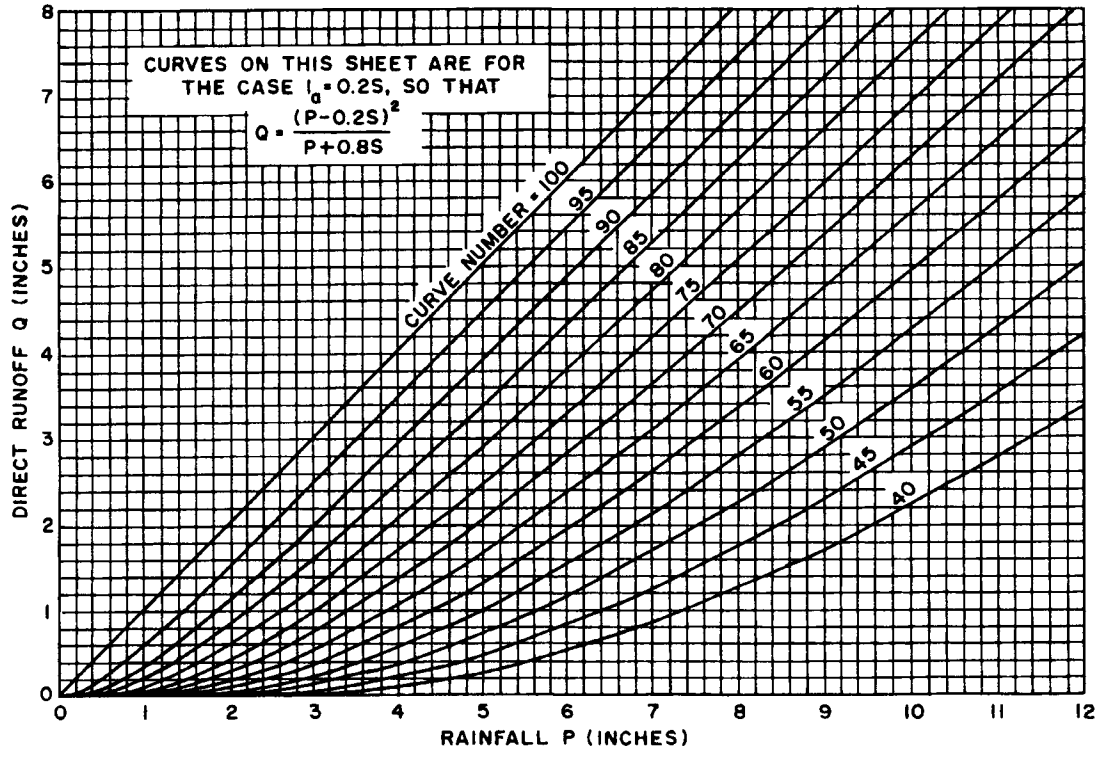


Figure 3.34.—Solution of Runoff Equation (SCS, 1972).

TABLE 3.4a. Runoff Curve Numbers for Urban Areas[1] (Soil Conservation Service, 1986).

Cover description		Curve numbers for hydrologic soil group—			
Cover type and hydrologic condition	Average percent impervious area[2]	A	B	C	D
Fully developed urban areas (vegetation established)					
Open space (lawns, parks, golf courses, cemeteries, etc.)[3]:					
Poor condition (grass cover < 50%)		68	79	86	89
Fair condition (grass cover 50% to 75%)		49	69	79	84
Good condition (grass cover > 75%)		39	61	74	80
Impervious areas:					
Paved parking lots, roofs, driveways, etc. (excluding right-of-way)		98	98	98	98
Streets and roads:					
Paved; curbs and storm sewers (excluding right-of-way)		98	98	98	98
Paved; open ditches (including right-of-way)		83	89	92	93
Gravel (including right-of-way)		76	85	89	91
Dirt (including right-of-way)		72	82	87	89
Western desert urban areas:					
Natural desert landscaping (pervious areas only)[4]		63	77	85	88
Artificial desert landscaping (impervious weed barrier, desert shrub with 1- to 2-inch sand or gravel mulch and basin borders)		96	96	96	96
Urban districts:					
Commercial and business	85	89	92	94	95
Industrial	72	81	88	91	93
Residential districts by average lot size:					
1/8 acre or less (town houses)	65	77	85	90	92
1/4 acre	38	61	75	83	87
1/3 acre	30	57	72	81	86
1/2 acre	25	54	70	80	85
1 acre	20	51	68	79	84
2 acres	12	46	65	77	82
Developing urban areas					
Newly graded areas (pervious areas only, no vegetation)[5]		77	86	91	94
Idle lands (CN's are determined using cover types similar to those in table 2-2c).					

[1]Average runoff condition, and $I_a = 0.2S$.

[2]The average percent impervious area shown was used to develop the composite CN's. Other assumptions are as follows: impervious areas are directly connected to the drainage system, impervious areas have a CN of 98, and pervious areas are considered equivalent to open space in good hydrologic condition. CN's for other combinations of conditions may be computed using Fig. 3.35 or 3.36.

[3]CN's shown are equivalent to those of pasture. Composite CN's may be computed for other combinations of open space cover type.

[4]Composite CN's for natural desert landscaping should be computed using figures 3.35/3.36 based on the impervious area percentage (CN = 98) and the pervious area CN. The pervious area CN's are assumed equivalent to desert shrub in poor hydrologic condition.

[5]Composite CN's to use for the design of temporary measures during grading and construction should be computed using Fig. 3.35 or 3.36 based on the degree of development (impervious area percentage) and the CN's for the newly graded pervious areas.

textures. The USDA soil textures normally included in this group are silt loam and loam. These soils have a transmission rate between 0.15 to 0.3 in/hr.

Group C soils have low infiltration rates when thoroughly wetted and consist chiefly of soils with a layer that impedes downward movement of water and soils with moderately fine to fine texture. The USDA soil texture normally included in this group is sandy clay loam. This soil has a transmission rate between 0.05 to 0.15 in/hr.

Group D soils have high runoff potential. They have very low infiltration rates when thoroughly wetted and consist mainly of clay soils with a high swelling potential, soils with a permanent high water

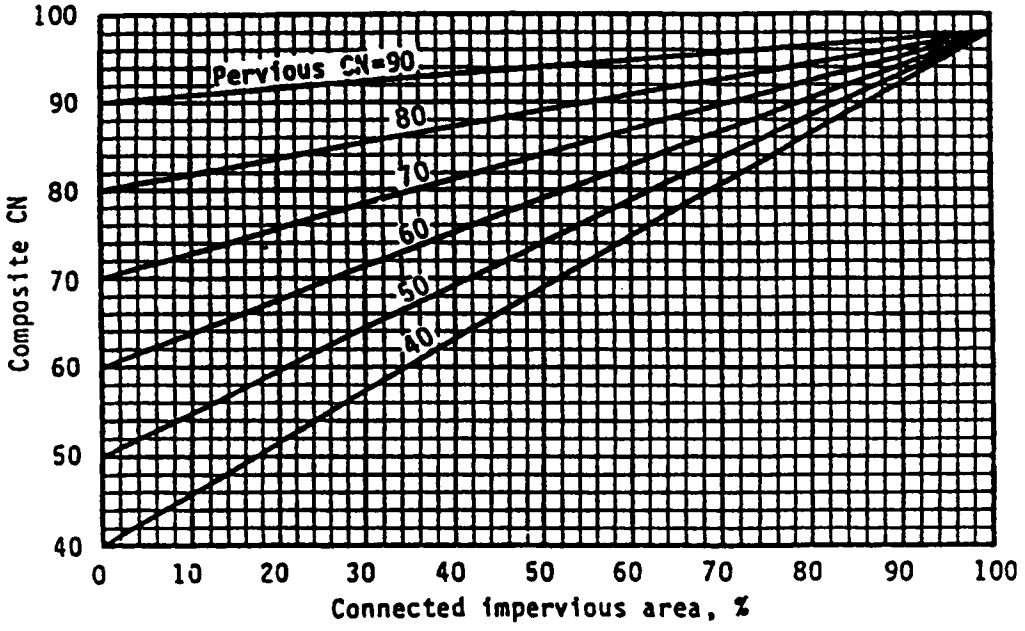


Figure 3.35.—Composite CN with Inconnected Impervious Areas and Total Impervious Areas Less Than 30%. (SCS 1986)

table, soils with a claypan or clay layer at or near the surface and shallow soils over a nearly impervious material. The USDA soil textures normally included in this group are clay loam, silty clay loam, sandy clay, silty clay, and clay. These soils have a very low rate of water transmission (0.0 to 0.05 in/hr). Some soils are classified in group D because of a high water table that creates a drainage problem; however, once these soils are effectively drained, the soils are placed into another group.

A list of most of the soils in the United States and their respective hydrologic soil group classification is given in Soil Conservation Service (1982a). Maps and soil reports are available on a county basis for most of the country and can be obtained from the library or SCS offices.

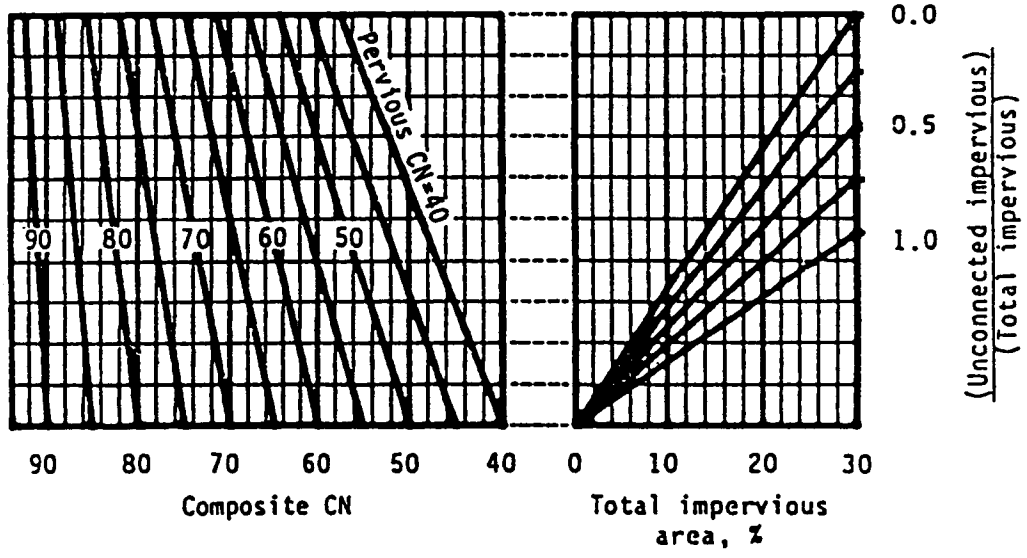


Figure 3.36.—Composite CN with Inconnected Impervious Areas and Total Impervious Areas Less Than 30%. (SCS 1986)

TABLE 3.4b. Runoff Curve Numbers for Cultivated Agricultural Lands[1] (Soil Conservation Service, 1986).

Cover description			Curve numbers for hydrologic soil group—			
Cover type	Treatment[2]	Hydrologic condition[3]	A	B	C	D
Fallow	Bare soil	—	77	86	91	94
	Crop residue cover (CR)	Poor	76	85	90	93
		Good	74	83	88	90
Row crops	Straight row (SR)	Poor	72	81	88	91
		Good	67	78	85	89
	SR + CR	Poor	71	80	87	90
		Good	64	75	82	85
	Contoured (C)	Poor	70	79	84	88
		Good	65	75	82	86
	C + CR	Poor	69	78	83	87
		Good	64	74	81	85
	Contoured & terraced (C&T)	Poor	66	74	80	82
		Good	62	71	78	81
	C&T + CR	Poor	65	73	79	81
		Good	61	70	77	80
Small grain	SR	Poor	65	76	84	88
		Good	63	75	83	87
	SR + CR	Poor	64	75	83	86
		Good	60	72	80	84
	C	Poor	63	74	82	85
		Good	61	73	81	84
	C + CR	Poor	62	73	81	84
		Good	60	72	80	83
	C&T	Poor	61	72	79	82
		Good	59	70	78	81
	C&T + CR	Poor	60	71	78	81
		Good	58	69	77	80
Close-seeded or broadcast legumes or rotation meadow	SR	Poor	66	77	85	89
		Good	58	72	81	85
	C	Poor	64	75	83	85
		Good	55	69	78	83
	C&T	Poor	63	73	80	83
		Good	51	67	76	80

[1]Average runoff condition, and $I_a = 0.2S$.
[2]*Crop residue cover* applies only if residue is on at least 5% of the surface throughout the year.
[3]Hydrologic condition is based on combination of factors that affect infiltration and runoff, including (a) density and canopy of vegetative areas, (b) amount of year-round cover, (c) amount of grass or close-seeded legumes in rotations, (d) percent of residue cover on the land surface (good ≥ 20%), and (e) degree of surface roughness.
Poor: Factors impair infiltration and tend to increase runoff.
Good: Factors encourage average and better than average infiltration and tend to decrease runoff.

b. Treatment Treatment is a cover-type modifier used in Table 3.4 to describe the management of cultivated agricultural lands. It includes mechanical practices, such as contouring and terracing, and management practices, such as crop rotations and reduced or no tillage.

c. Hydrologic Condition Hydrologic condition indicates the effects of cover type and treatment on infiltration and runoff and is generally estimated from density of plant and residue cover on sample areas. A good hydrologic condition indicates that the soil usually has a low runoff potential for the given hydrologic soil group, cover type, and treatment. Some factors to consider when estimating the effect of cover on infiltration and runoff are: 1) canopy or density of lawns, crops, or other vegetative areas; 2) amount of year-round cover; 3) amount of grass or close-seeded legumes in rotations; 4) percent of residue cover; and 5) degree of surface roughness. Several factors, such as the percentage of impervious

TABLE 3.4c. Runoff Curve Numbers for Other Agricultural Lands[1] (Soil Conservation Service 1986).

Cover description		Curve numbers for hydrologic soil group—			
Cover type	Hydrologic condition	A	B	C	D
Pasture, grassland, or range—continuous forage for grazing.[2]	Poor	68	79	86	89
	Fair	49	69	79	84
	Good	39	61	74	80
Meadow—continuous grass, protected from grazing and generally mowed for hay.	—	30	58	71	78
Brush—brush-weed-grass mixture with brush the major element.[3]	Poor	48	67	77	83
	Fair	35	56	70	77
	Good	[4]30	48	65	73
Woods—grass combination (orchard or tree farm).[5]	Poor	57	73	82	86
	Fair	43	65	76	82
	Good	32	58	72	79
Woods.[6]	Poor	45	66	77	83
	Fair	36	60	73	79
	Good	[4]30	55	70	77
Farmsteads—buildings, lanes, driveways, and surrounding lots.	—	59	74	82	86

[1]Average runoff condition, and $I_a = 0.2S$.
[2]*Poor*: < 50% ground cover or heavily grazed with no mulch.
Fair: 50 to 75% ground cover and not heavily grazed.
Good: > 75% ground cover and lightly or only occasionally grazed.
[3]*Poor*: < 50% ground cover.
Fair: 50 to 75% ground cover.
Good: > 75% ground cover.
[4]Actual curve number is less than 30; use CN = 30 for runoff computations.
[5]CN's shown were computed for areas with 50% woods and 50% grass (pasture) cover. Other combinations of conditions may be computed from the CN's for woods and pasture.
[6]*Poor*: Forest litter, small trees, and brush are destroyed by heavy grazing or regular burning.
Fair: Woods are grazed but not burned, and some forest litter covers the soil.
Good: Woods are protected from grazing, and litter and brush adequately cover the soil.

area and the means of conveying runoff from impervious areas to the drainage system, should be considered in computing CN for urban areas.

d. Antecedent Runoff Condition Antecedent runoff condition (ARC) is an index of runoff potential for a storm event. The ARC is an attempt to account for the variation in CN at a site from storm to storm. CN for the average ARC at a site is the median value as taken from sample rainfall and runoff data. The CN's in Table 3.4 are for the average ARC, which is used primarily for design applications. Soil Conservation Service (1982a) and Rallison (1980) give a detailed discussion of storm-to-storm variation and a demonstration of upper and lower enveloping curves.

e. Curve Number Limitations Curve numbers describe average conditions that are useful for design purposes. If the rainfall event used is a historical storm that departs from average conditions, the modeling accuracy decreases. The runoff curve number equation should be applied with caution when recreating specific features of an actual storm. The equation does not contain an expression for time. Therefore, it does not account for rainfall duration or intensity although Equation 3.11 can be applied to the cumulative rainfall at a number of points within the cumulative rainfall hyetograph. Thus, an excess rainfall hyetograph can be generated for the storm.

The user should understand the assumptions reflected in the initial abstraction term I_a and should ascertain that these assumptions apply to the situation at hand. I_a, which consists of interception, initial infiltration, surface depression storage, evapotranspiration, and other factors, is generalized as 0.2S based

Chapter 4

EVAPORATION AND TRANSPIRATION

I. INTRODUCTION

Evaporation from soil and water surfaces and transpiration from growing plants are major processes in the primary component of the hydrologic cycle that returns precipitated water to the earth's atmosphere as vapor. The primary force driving evaporation and transpiration is energy input from the sun. Transpiration is a special case of evaporation. Evaporation occurs at the surface of moist cells within plant tissues, and the water vapor then diffuses into intercellular leaf spaces and diffuses through stomates to the atmosphere. As a result, plants have some direct control of the process in contrast to evaporation from water surfaces. Other factors that effect evaporation from free water surfaces and transpiration from vegetated surfaces involve differences in reflectance of radiation, which affects radiant energy input, differences in heat storage capacities, and the aerodynamic roughness of water and vegetation which affects the transfer of sensible and latent heat.

The purpose of this chapter is to present current technology for estimating evaporation (E) from water bodies and evapotranspiration (ET) from soil and plants and to summarize principles and methodology for use by engineers involved in solving E and ET problems and in conducting E and ET studies. Evaporation and transpiration have been studied for centuries. For reviews of the history on the subject, see Brutsaert (1982) and Jensen et al. (1990).

Major progress in linking the processes with energy exchange came about in the twentieth century. The classic work of Penman (1948) laid the foundation for relating evapotranspiration to meteorological variables. Penman combined the energy balance component required to sustain evaporation with a mechanism to remove water vapor (sink strength). Many investigators, including Penman, continued to expand the theory of the combination equation since 1950 with special emphasis on the aerodynamic aspects. A detailed summary of the work in the 1950's and early 1960's was presented by Rijtema (1965) and Monteith (1965). Today, the most basic and practical equation for estimating evapotranspiration is commonly known as the "Penman-Monteith" method.

II. PHYSICS AND THEORY OF EVAPORATION

A. Surface–Air Energy Exchanges

1. Physical Properties of Water and Air Physical properties of liquid water and water vapor along with basic terminology are summarized in this section. Some important water properties are shown in Table 4.1.

Saturation vapor pressure, e^o, is an important parameter in hydrology. At the water–air interface there is a continuous flow of molecules from the water surface to the air, and a return flow to the liquid surface.

TABLE 4.1. Physical Properties of Liquid Water (Van Wijk and de Vries, 1963).

Temperature (°C)	Density (kg m^{-3})	Surface Tension (N m^{-1})	Dynamic Viscosity (10^{-3} Pa s)	Heat of vaporization (MJ kg^{-1})	Specific heat (kJ $kg^{-1}°C^{-1}$)	Thermal conductivity (J m^{-1} $s^{-1}°C^{-1}$)
−10	997.94	—	—	2.525	4.271	—
−5	999.18	0.0764	—	—	—	—
0	999.87	.0756	1.792	2.501	4.218	0.561
4	1000.00	.0750	—	2.492	4.205	.596
5	999.99	.0748	1.519	2.489	4.202	.574
10	999.73	.0742	1.308	2.477	4.192	.586
15	999.13	.0734	1.140	2.466	4.186	.595
20	998.23	.0727	1.005	2.453	4.182	.603
25	997.08	.0719	.894	2.442	4.180	.611
30	995.68	.0711	.801	2.430	4.178	.620
35	994.06	.0703	.723	2.418	4.178	.628
40	992.25	.0695	.656	2.406	4.178	.632
45	990.24	.0687	.599	2.394	4.179	.641
50	988.07	.0679	.549	2.382	4.181	.645

When equilibrium with pure water exists, the two flows are equal and the air is saturated with water vapor. The partial pressure exerted by the vapor at this time is called the saturation vapor pressure. The vapor pressure at equilibrium depends on the liquid water pressure, temperature, and its chemical content (solutes). Typical values of saturation vapor pressure and density of water vapor over a plane surface of pure water at the same temperature and pressure are summarized in Table 4.2. Saturation vapor pressure in kPa can be calculated following Tetens (1930) and Murray (1967) for T in °C as:

$$e^o = \exp\left(\frac{16.78\,T - 116.9}{T + 237.3}\right). \tag{4.1}$$

This expression calculates within 0.1% of values in the Smithsonian Meteorological Tables for temperatures in the range of 0 to 50°C (Allen et al., 1989). The slope of the saturation vapor pressure curve, Δ, kPa $°C^{-1}$, is obtained by differentiating Equation 4.1 (Allen et al., 1989):

$$\Delta = \frac{4098\,e^o}{(T + 237.3)^2}. \tag{4.2}$$

Absolute humidity, ρ_v, is the water vapor density, i.e., the mass of water vapor per unit volume of moist air. It can be calculated from the ideal gas law:

$$\rho_v = \frac{P_v}{R_v T}, \tag{4.3}$$

where ρ_v is the absolute humidity, kg m^{-3}, P_v is the vapor pressure, N m^{-2} or Pa, R_v is the gas constant for water vapor (461.5 J kg^{-1} K^{-1}), and T is the absolute temperature in K (K = 273.2 + °C). For ρ_v in kPa the density in kg m^{-3} is:

$$\rho_v = \frac{2.167\,P_v}{T}. \tag{4.4}$$

Specific humidity, q, is the relative density of water vapor, i.e., the mass of water vapor per unit mass of moist air:

TABLE 4.2. Physical Properties of Water Vapor (List, 1984; ASTM, 1976).

Temperature (°C)	Saturation vapor pressure over water (kPa)	Saturation vapor density over water (10^{-3} kg m^{-3})	Diffusion coefficient at P = 100 kPa (10^{-5} m^2 s^{-1})
0	0.611	4.85	2.26
5	.872	6.80	—
10	1.227	9.40	2.41
15	1.704	12.83	—
20	2.337	17.30	2.57
25	3.167	23.05	—
30	4.243	30.38	2.73
35	5.624	39.63	—
40	7.378	51.19	2.89
45	9.586	65.50	—
50	12.340	83.06	—

1 kPa = 10 mb = 7.501 mm Hg
10^{-3} kg m^{-3} = 10^{-6} g cm^{-3}

$$q = \frac{\rho_v}{\rho_a} = \frac{m_v}{m_v + m_a} = \frac{0.622\ e}{P - 0.378\ e}, \tag{4.5}$$

where q is the specific humidity, kg kg^{-1}, ρ_a is the density of moist air, m_v is the mass of water vapor, m_a is the mass of dry air, P is the total atmospheric pressure in kPa, e is the vapor pressure in kPa, and 0.622 is the ratio of the molecular mass of water to the apparent molecular mass of dry air.

Mixing ratio, r, is the dimensionless ratio of the mass of water vapor to a unit dry air mass. At the same temperature, and when e is small relative to the total pressure P, the mixing ratio is:

$$r = \frac{m_v}{m_a} \simeq 0.622\ \frac{e}{P}. \tag{4.6}$$

A more precise equation (than Equation 4.1) for calculating saturation vapor pressure, e^o, of air over water at pressure P and temperature is:

$$e^o = \frac{r^o}{0.622 + r^o} P, \tag{4.7}$$

where r^o is the saturation mixing ratio.

Saturation deficit, d_a, or vapor pressure deficit of the air is:

$$d_a = e^o - e. \tag{4.8}$$

Relative humidity, RH, is the dimensionless ratio of actual vapor pressure to saturation vapor pressure, usually expressed in percent. RH and d_a are defined as:

$$RH = 100\ \frac{e}{e^o} \tag{4.9}$$

and

$$d_a = e^o\left[\frac{100 - RH}{100}\right]. \tag{4.10}$$

Relative humidity by itself has limited utility in E and ET calculations unless air temperature or one of the vapor terms corresponding to the same time scale and interval as RH is given.

Dew point temperature is the temperature to which a parcel of air must be cooled at constant pressure and constant vapor content until saturation occurs, or the temperature at which saturation vapor pressure is equal to the actual vapor pressure of the contained vapor, and r is the mixing ratio defined in Equation 4.6. At dew point temperature, condensation normally occurs:

$$e_d^o = \frac{r}{0.622 + r} P. \tag{4.11}$$

Wet bulb temperature, T_w, is the temperature that a moist evaporating surface may approach when the radiation energy balance is zero. In practice, the wet bulb temperature represents the equilibrium temperature of a thermometer covered with a cloth that has been wetted with pure water in air moving at least 4.6 m/s. The wet bulb thermometer is cooled until heat drawn from the air equals the gain of latent heat due to evaporation. The vapor pressure of the ambient air is:

$$e = e_w^o - \frac{Pc_p}{0.622\ \lambda}(T - T_w) = e_w^o - \gamma(T - T_w), \tag{4.12}$$

where e_w^o is saturation vapor pressure at T_w, P is atmospheric pressure, c_p is the specific heat of air at constant pressure, λ is the latent heat of vaporization, T is ambient air temperature, and γ is referred to as the psychrometric constant for fully aspirated psychrometers. At sea level, normal values are: P (101.3 kPa at one atm.), c_p (1.003×10^{-3} MJ kg^{-1} °C^{-1}), and λ (2.453 MJ kg^{-1} at 20 °C). The pyschrometric constant, γ, is computed as:

$$\gamma = \frac{c_p P}{\lambda \epsilon}, \tag{4.13}$$

where, for $c_p = 1.013$ kJ kg^{-1} K^{-1} for moist air (Brutsaert, 1982), P is in kPa, λ is in kJ kg^{-1}, and $\epsilon = 0.622$. γ has units of kPa °C^{-1}. Mean atmospheric air pressure, P, can be computed using the ideal gas law (Burman et al., 1987) as:

$$P = P_o\left(\frac{T_o - \alpha\,(z - z_o)}{T_o}\right)^{\frac{g}{\alpha_a R}}, \tag{4.14}$$

where P_o and T_o are known atmospheric pressure in kPa and absolute temperature in K at elevation z_o in m, and z is the elevation of the location or instrument in m above mean sea level. The assumed constant adiabatic lapse rate, α_a, normally is taken as 0.0065 K m^{-1} for saturated air or 0.01 K m^{-1} for non-saturated air; g is gravitational acceleration, 9.81 m/s and R is the specific gas constant for dry air, 287.0 J kg^{-1} K^{-1}. Values for P_o, T_o, and z_o are commonly those for the standard atmosphere at sea level, which are 101.3 kPa, 288 K, and 0 m, respectively (List, 1984). Equation 4.14 is relatively insensitive to the value of α_a for elevations up to 3,000 m. Air density, ρ_a, can be computed as

$$\rho_a = \frac{1000\ P}{T_v R} \tag{4.15}$$

where P is in kPa, R is 287.0 J kg^{-1} K^{-1}, and T_v is virtual temperature in K. T_v can be computed as follows (Jensen et al., 1990):

$$T_v = \frac{T}{1 - 0.378\frac{e}{P}} \tag{4.16}$$

where T is air temperature, K, and e is mean vapor pressure of the air, kPa.

Latent heat of vaporization, λ, in MJ kg^{-1} can be computed following Harrison (1963) as:

$$\lambda = 2.501 - 2.361 \times 10^{-3}\ T, \tag{4.17}$$

where T is mean air temperature in °C. If available, mean surface temperature or wet bulb temperature can be used to compute the value of λ, which better represents conditions at the evaporating surface.

2. Standard Atmosphere (Lower Atmosphere) Characteristics of a standard atmosphere (average conditions for the United States at latitude 40°N) are given in Table 4.3. Characteristics of standard atmospheres at other latitudes are summarized by Burman et al. (1987).

B. Radiation Balance

Historical and recent studies of evaporation (E) and evapotranspiration (ET) clearly show, where soil water is not limiting, that the primary variable controlling the rate of E and ET is solar radiation impinging on the evaporating surfaces. Part of the solar radiation, R_s, is reflected back to the atmosphere and part of that which is absorbed by the surface is reradiated back as long-wave radiation, although this loss of energy is compensated in part by the downcoming long-wave from the sky. The radiation balance determines the net radiant energy, R_n, available at the evaporating surfaces. The most accurate methods of estimating E and ET require determining the radiation balance, or net radiation for the surface. Many studies have shown that daily net radiation is closely related to the daily rate of E from shallow water bodies and ET when soil water is not limiting, especially in warm to hot subhumid and humid climates. During daytime, the responses of E or of ET to changes in solar and net radiation are closely linked even over periods as short as five minutes. Examples in the next section reveal that the actual fraction of R_n going into E or ET can vary markedly depending upon air mass conditions above the surface.

1. Solar Radiation The principal source of heat energy for E and ET is solar radiation, R_s. When measured at the earth's surface, R_s includes both direct and diffuse short-wave radiation and may be called "global radiation." The most recent measurement of the solar constant, or the rate at which solar radiation is received on a surface normal to the incident radiation outside the earth's atmosphere is 1.367 kJ m^{-2} s^{-1} (London and Frohlich, 1982; Lean, 1989).

On cloudless days, the atmosphere is relatively transparent to solar or short wave radiation. About 70 to 80 percent of extraterrestrial radiation reaches the earth's surface in semiarid areas. The balance is reflected from dust particles or is absorbed by various gases in the atmosphere. Techniques for estimating atmospheric transmittance are presented by Idso (1969; 1970; 1981) and Davies and Idso (1979). Solar

TABLE 4.3. Standard Lower Atmosphere (Adapted from List, 1984).

Altitude (m)	Temperature (°C)	Pressure (kPa)	Pressure (mb)	Density (kg m^{-3})
0	15.0	101.3	1013	1.226
500	11.75	95.5	955	1.168
1,000	8.50	89.9	899	1.112
1,500	5.25	84.6	846	1.058
2,000	2.00	79.5	795	1.007
3,000	−4.50	70.1	701	.909

radiation can be measured directly with instruments ranging from a spring-wound chart and bi-metallic sensors to complex sensors and automated data-logging systems.

Estimates of daily cloudless day solar radiation received at the earth's surface, R_{so}, can be made using extraterrestrial radiation, R_A, and atmospheric transmittance. Clear sky short-wave radiation, R_{so}, can be approximated under most conditions as

$$R_{so} = k_A \, R_A, \tag{4.18}$$

where k_A is a clearness index which is largely a function of station elevation (atmospheric thickness) and atmospheric turbidity. For conditions of low turbidity, k_A can be computed as a function of station elevation for locations less than 6,000 m using a relationship developed by Allen et al. (1994b):

$$k_A = 0.75 + 0.00002 \, E, \tag{4.19}$$

where E is elevation in m. This relationship was obtained by integrating Equation 4.15 (air density) with respect to elevation and assuming constant turbidity with elevation, absorption, and scattering of solar radiation in proportion to air density. Atmospheric pressure in Equation 4.15 was computed using Equation 4.14. Equation 4.19 predicts $k_A = 0.75$ at sea level, which agrees with Equation 4.27 for $n/N = 1$. Equation 4.19 also agrees with long-term measurements by Wright (1978) at Kimberly, Idaho, elevation 1,200 m, where $R_{so}/R_A = 0.78$. For areas of high turbidity due to pollution or airborne dust, k_A may need to be reduced by up to 10% depending on the relative turbidity.

Estimates of R_{so} for a given area can also be developed from an envelope curve plotted through measured radiation data on cloudless days. The relationship $R_{so} = k_A \, R_A$ is useful for verifying correct operation of pyranometers. Ordinarily, R_{so} should plot as an upper envelope of measured R_s. Reflection of solar radiation by adjacent clouds can occasionally increase measured R_s to levels which are above R_{so}.

Extraterrestrial radiation, R_A, can be computed for a location as a function of latitude and day of the year using the following mathematical equations from Duffie and Beckman (1991) (Jensen et al., 1990):

$$R_A = \frac{24(60)}{\pi} G_{sc} \, d_r [\omega_s \sin(\phi) \sin(\delta) + \cos(\phi) \cos(\delta) \sin(\omega_s)] \tag{4.20}$$

where R_A is daily extraterrestrial radiation in MJ m^{-2} d^{-1} and ϕ is latitude of the station in radians (negative for southern latitudes). The declination, δ, in radians can be estimated as:

$$\delta = 0.4093 \sin\left(2\pi \frac{(284 + J)}{365}\right), \tag{4.21}$$

where J is the day of the year (January 1st = 1). The term d_r is the relative distance of the earth from the sun, where:

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi}{365} J\right). \tag{4.22}$$

The sunset hour angle, ω_s, in radians can be calculated using Equation 4.23 or 4.24:

$$\omega_s = \arccos[-\tan(\phi) \tan(\delta)] \tag{4.23}$$

$$\omega_s = \frac{\pi}{2} - \arctan\left[\frac{-\tan(\phi) \tan(\delta)}{[1 - \tan^2(\phi) \tan^2(\delta)]^{1/2}}\right]. \tag{4.24}$$

G_{sc} in the R_A equation is the solar constant of 0.0820 MJ m^{-2} min^{-1} (1367 W m^{-2}) as recommended by the International Association of Meteorology and Atmospheric Physics (London and Frohlich, 1982). The $-\tan(\phi)\tan(\delta)$ expression in Equation 4.23 must be limited to less than or equal to 2.0 in extreme latitudes (> 55°) during winter months. If the $-\tan(\phi)\tan(\delta)$ expression is less than -1.0 in extreme latitudes (> 55°) during summer months, then the $-\tan(\phi)\tan(\delta)$ expression should be set equal to $[\tan(\phi)\tan(\delta) - 2.0]$. Equations for estimating hourly extraterrestrial and clear-sky radiation can be found in Allen (1996) and Hottel (1976) as summarized in Appendix B of ASCE Manual 70 (Jensen et al., 1990). Heermann et al. (1985) developed an empirical equation for estimating daily R_{so} in the United States; however, the Heermann equation is applicable only to northern latitudes between 25 and 65 degrees during growing season periods and is therefore not presented here. Equation 4.18 is recommended instead.

Solar radiation, R_s, on a given day is affected mainly by cloud cover. Therefore, R_s can be estimated using R_{so} and either degree of cloud cover or percent of sunshine. More reliable R_s estimates are obtained using recorded percent sunshine as compared to observed cloud cover data because observed cloud cover data are qualitative. Constants developed for a linear equation by Fritz and MacDonald (1949) are similar to those obtained in Canada and Australia:

$$R_s = (0.35 + 0.61S)R_{so}, \tag{4.25}$$

where S represents the ratio of actual to possible sunshine. Equation 4.25 allows for the fact that no recording instrument gives a full record of sunshine even on cloudless days. Where cloud cover is not strongly influenced by local orographic features, measurement of daily R_s at a single station may be used for estimates over large areas (100–200 km in diameter) without significant error over 5- or 10-day periods.

Solar radiation also can be estimated from extraterrestrial radiation, R_A. Black et al. (1954) correlated extraterrestrial solar radiation and duration of sunshine as recorded by Marvin and Campbell-Stokes sunshine recorders. The resulting equation based on 32 stations is:

$$R_s = (0.23 + 0.48S)R_A. \tag{4.26}$$

Doorenbos and Pruitt (1977) recommended a generalized form of the Penman R_s equation:

$$R_s = \left(0.25 + 0.5\,\frac{n}{N}\right) R_A. \tag{4.27}$$

The quantity n/N was defined by Doorenbos and Pruitt (1977) as "the ratio between actual measured bright sunshine hours and maximum possible sunshine hours." In practice, n/N and S are usually assumed to be the same quantity. Doorenbos and Pruitt presented a table to convert cloudiness expressed in eighths (octas) or in tenths to n/N values. Experimentally-derived constants for Equation 4.27 are presented in an appendix table for a number of specific locations by Doorenbos and Pruitt (1977). Variable N in Equation 4.27 can be calculated from the sunset hour angle as:

$$N = \omega_s\left(\frac{24}{\pi}\right) \tag{4.28}$$

where ω_s is sunset hour angle in radians, calculated using Equation 4.23 or 4.24.

Hargreaves and Samani (1982) proposed estimating R_s from the range in daily air temperature:

$$R_s = K_{RS}\,(T_x - T_n)^{0.5}\,R_A \tag{4.29}$$

where T_x and T_n are maximum and minimum daily air temperature, °C. Variable K_{RS} is an empirical coefficient equal to about 0.16 for interior regions (Hargreaves and Samani, 1982) and about 0.19 for

coastal regions (Hargreaves, 1994). Equation 4.29 performs best using mean monthly data. When applied to daily data, Equation 4.29 tends to overestimate for cloudy days. R_s predicted by Equation 4.29 should be limited to $\leq R_{so}$. Bristow and Campbell (1984) developed an equation similar to Equation 4.29; however, their formulation is more complex to apply and has similar accuracy.

2. Solar Radiation Data Base, United States The National Renewable Energy Laboratory (NREL) has compiled a mean monthly solar radiation data base and associated 30-year (1961-1990) record for the United States (NREL, 1992). These data include measured and modeled solar radiation and meteorological data for 239 locations in the United States and its possessions. They are available on magnetic media (disks and tape) from the National Climate Data Center in Asheville, North Carolina. Future releases will be produced on CD-ROM.

3. Net Radiation Net radiation requires the measurement or estimates of both incoming and reflected short wave radiation (< 4 μm) and net long-wave radiation. The atmosphere is much less transparent to long-wave radiation as compared to short-wave radiation. Water vapor, carbon dioxide, and ozone are good absorbers and emitters at some infrared wave lengths. Radiant energy absorbed by these various gases is emitted in all directions according to the radiation law.

$$R = \epsilon \sigma T^4, \tag{4.30}$$

where ϵ is emissivity, σ is the Stefan-Boltzmann constant, and T is the absolute temperature, K. Values of the Stefan-Boltzmann constant are $\sigma = 5.675 \times 10^{-8}\,\mathrm{J\,m^{-2}\,K^{-4}\,s^{-1}}$, or $\sigma = 4.903 \times 10^{-9}\,\mathrm{MJ\,m^{-2}\,K^{-4}\,d^{-1}}$.

Because of the high transparency of the atmosphere to the 8 to 13 μm wavelengths, there is a net loss of radiant energy to the atmosphere within this range during cloudless and partly cloudy 24-hour periods. For practical purposes, only the long-wave radiation components are important between sunset and sunrise. A small amount of diffuse short-wave sky radiation is involved just after sunset and prior to sunrise, respectively.

Net radiation flux can be measured directly using hemispherical net radiometers or estimated from net short-wave and net long-wave components.

$$R_n = R_s \downarrow - \alpha R_s \uparrow + R_L \downarrow - R_L \uparrow = (1 - \alpha) R_s \downarrow - R_b \uparrow, \tag{4.31}$$

where R_s is short-wave radiation or solar radiation, and R_L is long-wave radiation. The arrows indicate direction; $(1 - \alpha)R_s$ represents the net short-wave radiation received by a water, soil, or vegetated surface; α is the short-wave reflectance or albedo and R_b is the net back, or net outgoing, long-wave radiation.

Rather than estimating net radiation for each vegetated surface, or for each crop and growth stage, it is most common to measure or estimate the net radiation for a well-watered reference crop like grass or alfalfa. The resulting estimated rate of *ET* for other crops is then related to the reference crop. Mean daily reflectance coefficients (albedo) for most green field crops with a full cover ranges from 0.20 to 0.25. A value of 0.23 for α is commonly used for vegetated surfaces with a full cover. Full cover exists when the leaf area index, which is the leaf area per unit land area, is greater than about 3.

Wright (1982) presented an equation for estimating net radiation received by alfalfa in Kimberly, Idaho in which the albedo varies during the growing season (Jensen et al., 1990, Eq. 6.67, p. 137). The albedo values for inland water bodies range from 0.04 to 0.09. Bolsenga and Vanderploeg (1992) suggested a range from 0.07–0.15 depending upon water-surface conditions. Brutsaert (1982) summarized albedo values as reported earlier by a number of researchers (Table 4.4).

Long-wave radiation emitted by a surface can be estimated using Equation 4.30 if the surface temperature is known. The emissivity of a mixture of vegetation and soil surfaces, ϵ_{vs}, is about 0.98. Table 4.5 lists reported ϵ_{vs} values for a number of surfaces. The emissivity of water surfaces is about 0.97. The incoming long-wave radiation is more difficult to estimate since it must be an integrated value based on variations

TABLE 4.4. Approximate Mean Albedo Values For Various Natural Surfaces (Brutsaert, 1982).

Nature of surface	Albedo
Deep Water	0.04-0.08
Moist dark soils; plowed fields	0.05-0.15
Gray soils; bare fields	0.15-0.25
Dry soils; desert	0.20-0.35
White sand; lime	0.30-0.40
Green grass and other short vegetation (e.g., alfalfa, potatoes, beets)	0.15-0.25
Dry grass; stubble	0.15-0.20
Dry prairie and savannah	0.20-0.30
Coniferous forest	0.10-0.15
Deciduous forest	0.15-0.25
Forest with melting snow	0.20-0.30
Old and dirty snow cover	0.35-0.65
Clean, stable snow cover	0.60-0.75
Fresh dry snow	0.80-0.90

TABLE 4.5. Values of the Emissivities, ϵ, of Some Natural Surfaces.

Nature of surface	Emissivity
Bare soil (mineral)	0.95-0.97
Bare soil (organic)	0.97-0.98
Grassy vegetation	0.97-0.98
Tree vegetation	0.96-0.97
Snow (old)	0.97
Snow (fresh)	0.99

in water vapor and temperature, with elevation above the ground surface, degree of cloud cover and temperature of the clouds. It can be directly measured during dark hours with a total hemispheric radiometer (upward facing), or indirectly computed during daylight hours by subtracting R_s from the radiometer readings. Dense clouds act as a black body and essentially block the long-wave window (8 to 13 μm) to outer space and they emit long-wave radiation to the ground. The temperature at the cloud base is required to calculate this component of incoming thermal radiation.

Calculation of net thermal radiation becomes complex for partly cloudy conditions. A detailed analysis of the factors affecting net radiation involves both the atmospheric and surface emittances and cloud cover. Calculation of R_n on sloping terrain and related references is discussed in the section "Evapotranspiration from Land Surfaces—Direct Penman-Monteith."

Kimball et al. (1982) derived a complex model involving the areas of cloud cover and cloud temperature and transmittance in the 8 to 13 μm wavelength window. For a more complete discussion, see Davies and Idso (1979) and Dong et al. (1992).

Since measured incoming long-wave radiation data generally are not available, and temperature and vapor pressure as a function of elevation above the ground surface likewise are generally not available, most estimates of incoming long-wave radiation are based on air temperature and vapor pressure at instrument shelter height. The effects of cloud cover on long-wave atmospheric radiation and simple empirical equations to estimate net long-wave radiation are presented below.

The early work by Brunt (1932) relates clear-sky atmospheric emittance, ϵ_a, to vapor pressure at screen height, e_d, which takes the general form:

$$\epsilon_a = a_b + b_b\sqrt{e_d}, \tag{4.32}$$

where a_b and b_b are empirical coefficients.

For cloudy conditions, the net long-wave or thermal radiation, R_b, is influenced by the degree of cloud cover, its elevation and temperature.

$$R_b = R_L \uparrow - R_L \downarrow = R_{vs} \uparrow - (R_a + R_c)\downarrow, \tag{4.33}$$

where R_L is long-wave radiation, R_{vs} is long-wave radiation emitted by the vegetated or soil surface, R_a is full-spectrum long-wave radiation emitted by a cloudless sky and R_c is additional long-wave radiation emitted by clouds and transmitted to the earth's surface through the 8–13 μm wave band. In practice, the two long-wave radiation terms in Equation 4.33 are usually calculated using only instrument height data. Since air temperature at instrument height is used to estimate both incoming and outgoing long-wave radiation, then:

$$R_b = f\,(\epsilon_{vs} - \epsilon_a)\,\sigma T^4, \tag{4.34}$$

where ϵ_a is atmospheric emittance, ϵ_{vs} is vegetative and soil emittance, T is air temperature in K, and f is a factor to adjust for cloud cover. Equation 4.34 assumes that the vegetated surface temperature and air temperature at instrument height are equal.

The two emittances may be combined into a single term known as the net emittance, ϵ', which represents the difference between atmospheric and the combined vegetative and soil emittance when a single temperature is used. Values of ϵ_{vs} for common, fully developed crops range from 0.94 to 0.995. A canopy value of 0.98 is recommended for most radiation balance calculations over vegetated surfaces. A value of 0.97 is recommended for water.

Wright and Jensen (1972) developed the following expression for f in Equation 4.34 which represents an average of day-time and night-time conditions:

$$f = a\frac{R_s}{R_{so}} + b. \tag{4.35}$$

The Brunt form of the net emittance expression ϵ' is:

$$\epsilon' = \epsilon_{vs} - \epsilon_a = a_1 + b_1\sqrt{e_d}, \tag{4.36}$$

where e_d is the saturation vapor pressure at dewpoint temperature.

Combining Equations 4.34, 4.35, and 4.36 results in a general equation for net outgoing long-wave radiation.

$$R_b = \left(a\frac{R_s}{R_{so}} + b\right)\left(a_1 + b_1\sqrt{e_d}\right)\sigma T^4. \tag{4.37}$$

For 24-hour or longer time steps, T^4 in Equation 4.37 should be replaced by $(T_x^4 + T_n^4)/2$ for improved accuracy, where T_x and T_n are mean maximum and minimum daily air temperatures in K at screen or shelter height for the period. Experimental values of the coefficients for use in Equation 4.37 are presented in Table 4.6.

In practice, the net long wave radiation is first estimated for clear skies using air temperature and humidity data measured at screen height as follows:

$$R_{bo} = \epsilon'\sigma T^4 = \left(a_1 + b_1\sqrt{e_d}\right)(4.903 \times 10^{-9})\frac{(T_x^4 + T_n^4)}{2}, \tag{4.38}$$

where R_{bo} is net long-wave radiation for clear skies, ϵ' is the net emissivity when using only instrument height temperature, e_d is the saturation vapor pressure at mean dewpoint temperature in kPa, 4.903×10^{-9} is the Stefan-Boltzmann constant in MJ m^{-2} d^{-1} K^{-4} for units of R_{bo} in MJ m^{-2} d^{-1}, and T_x and T_n

TABLE 4.6. Experimental Coefficients for Net Long-Wave Radiation Equations 4.35 to 4.38 (Jensen et al., 1990).

Region	$(a,$	$b)$	$(a_1,$	$b_1)$
Davis, California	(1.35,	−0.35)[1]	(0.35,	−0.145)[2]
Southern Idaho	(1.22,	−0.18)[3]	(0.325,	−0.139)[3]
England	(not available)		(0.47,	−0.206)[4]
England	(not available)		(0.44,	−0.253)[5]
Australia	(not available)		(0.35,	−0.133)[6]
General	(1.3,	−0.3)[7]	(0.39,	−0.158)[8]

For e in mb, divide b_1 values by $\sqrt{10}$.
[1]Pruitt (1976).
[2]Goss and Brooks (1956).
[3]Wright and Jensen (1972).
[4]Monteith and Szeicz (1962).
[5]Penman (1948).
[6]Fitzpatrick and Stern (1965).
[7]Suggested for arid areas.
[8]Budyko (1956).

are in K. If Equation 4.38 is applied on an hourly or shorter time step, then T_x and T_n can be set equal to the mean temperature for the time step.

Since a single level temperature measurement is used, $(a_1 + b_1 \sqrt{e_d})$ represents ϵ' or the difference between the emittance for the reference crop and the effective emittance for the atmosphere. The dewpoint temperature does not change greatly during the day, and a single level dewpoint observation during the day is adequate for most estimates of R_{bo} for eventual calculation of E or ET.

When solar radiation data are available, they can be used to adjust the net long-wave radiation for clear skies or for partly cloudy conditions.

$$R_b = \left(a \frac{R_s}{R_{so}} + b \right) R_{bo}. \tag{4.39}$$

The coefficients a and b are derived empirically and $a + b = 1.0$. Smith et al. (1991) and Allen et al. (1994b) have recommended using $a = 1.35$, $b = -0.35$, $a_1 = 0.34$, and $b_1 = -0.14$ at most locations for consistency in standardized reference ET calculations.

Where humidity measurements are not available, minimum air temperatures may be taken as the dewpoint temperature in semihumid and humid areas. In arid areas, the dewpoint temperature may be several degrees less than the minimum air temperature. Simple calibration relationships can be developed to enable using minimum air temperature as a substitute for dewpoint temperature.

C. Energy Balance

In cold humid climates, or in cold winter months of temperate to semiarid zones, only 50 to 60 percent of the net radiation, R_n, may be converted to latent heat. In hot, arid climates, latent heat may exceed net radiation by 10 to 50 percent with sensible heat derived from the air and converted to latent heat. In spite of these relationships, the heat energy balance approach to determining or estimating E and ET is recognized as a reliable and conservative method. A thorough understanding of the factors controlling the energy balance of a water body or land area enables accurate estimates or predictions of E and ET.

Whereas the radiation balance considers only radiation terms in developing the net exchange of radiation (Equation 4.33), the energy balance relates to the various ways in which R_n is balanced by the inputs or outputs of energy from non-radiative parameters. The vertical energy balance at the soil or water surface or at the "effective surface" of a crop is the sum of sensible heat flux to or from the air and

soil (or of water), latent heat, net radiation, and other miscellaneous fluxes. The major components of the vertical energy balance can be expressed as follows:

$$R_n = \lambda E + H + G, \tag{4.40}$$

where R_n is the net radiation flux at the surface, λE is the latent heat flux (positive during evaporation), H is the sensible heat exchange from the surface to air (positive if the air is warming), and G is the sensible heat exchange from the surface to the soil or water (positive if the soil or water is warming). The miscellaneous terms such as heat storage within the foliage, photosynthesis, and respiration are generally insignificant relative to magnitudes of R_n, λE, H, and G, and are neglected in Equation 4.40. For daily or longer periods, G can be ignored, although with significant variation of R_n and/or air mass conditions from one day to the next, this may not be advisable. For shorter periods, e.g., hourly, G must be taken into account. Some of the minor terms such as changes in heat storage in a canopy may also need attention, especially in forest canopies. If R_n and sensible heat flux densities H and G can be measured or estimated reliably, then the latent heat flux density, λE, can be computed from Equation 4.40.

Calculation of R_n has been discussed earlier in this section. Description and equations for calculating H and λE are presented later in the sections on Energy Balance and Mass Transfer Methods. Calculation of G is discussed in the following section.

D. Sensible Heat Flux Density—Soil

The magnitude of soil heat storage or release can be significant over a few hours but is usually small from day to day because heat stored early in the day as the soil warms is lost late in the day or at night when the soil cools. The rate of heat storage or release at any depth z can be expressed by:

$$G = K_{Ts} \frac{\partial T_s}{\partial z}, \tag{4.41}$$

where T_s is the temperature of the soil and K_{Ts} is the thermal conductivity of the soil in J m^{-1} s^{-1} $°C^{-1}$. Therefore, G in Equation 4.41 has units of J m^{-2} s^{-1} (W m^{-2}) for T_s in °C and z in m. Typical values of K_{Ts} are presented in Table 4.7. K_{Ts} for soils varies widely with soil moisture content. The effect of soil moisture is to provide thermal continuity between soil particles, thereby increasing bulk K_{Ts}. Brutsaert (1982,

TABLE 4.7. Thermal Properties of Soil Constituents at 20°C and Standard Atmospheric Pressure (van Wijk and de Vries, 1963).

Soil material	Density heat ρ [Mg m^{-3}]	Specific heat [kJ $kg^{-1}C^{-1}$]	Volumetric heat capacity C_s [MJ m^{-3} C^{-1}]	Thermal conductivity λ [J m^{-1} s^{-1} C^{-1}]	Thermal diffusivity α [10^{-6} m^2 s^{-1}]
Quartz	2.65	0.733	1.93	8.37	4.3
Minerals*	2.65	0.733	1.93	2.93	1.5
Organic matter*	1.3	1.926	2.51	0.25	0.1
Water	1.00	4.187	4.19	0.59	0.142
Air	0.0012	1.005	0.0012	0.026	0.021

*Approximate average values
1000 kg m^{-3} = 1 g cm^{-3}
4186.8 J kg^{-1} C^{-1} = 1 cal g^{-1} C^{-1}
4.1868 MJ m^{-3} C^{-1} = 1 cal cm^{-3} C^{-1}
0.41 868 J m^{-1} s^{-1} C^{-1} = 0.41 868 W m^{-1} C^{-1} = 10^{-3} cal cm^{-1} s^{-1} C^{-1}
10^{-6} m^2 s^{-1} = 10^{-2} cm^2 s^{-1}

Figure 6.9) presented curves of thermal conductivity as a function of volumetric water content for four widely different soils. These values ranged from 0.1 to 2.5 J m^{-1} s^{-1} K^{-1}.

Soil or water heat flux density can also be estimated by monitoring the change in temperature of a soil or water profile over time:

$$G = C_s \int_0^{z_s} \frac{\partial T_s}{\partial t} dz, \tag{4.42}$$

where C_s is the heat capacity per unit volume of soil, MJ m^{-3} $°C^{-1}$, t is time, and z_s is the depth of soil which responds to increased or decreased temperature change. C_s can be estimated from the following equation by de Vries (1963):

$$C_s = 1.93\ V_m + 2.51\ V_{om} + 4.19\ \theta, \tag{4.43}$$

where V_m, V_{om}, and θ represent the volume of minerals, organic matter, and water per unit volume of soil, respectively. For example, if a soil has 50% solids and negligible organic matter and a volumetric water content of 0.27, its heat capacity will be 2.1 MJ m^{-3} $°C^{-1}$. Approximate values for soils and soil components are given in Table 4.7. Utilization of Equation 4.42 in combination with soil heat flux plates is common.

Since the magnitude of daily soil heat flux under a crop canopy over 10- to 30-day periods or longer is relatively small, the daily value of G normally can be neglected for most practical estimates involving energy balance; however, the total value of G over the entire period may be significant, especially for 30 days or longer.

An approximation of the soil heat flux involved over long periods of time ($\geq$30 days) can be obtained by assuming that the soil temperature to a depth of 2 m changes approximately with average air temperature and that the average volumetric heat capacity for the soil is 2.1 MJ m^{-3} $°C^{-1}$,

$$G = -4.2 \frac{T_{i-1} - T_{i+1}}{\Delta t} = 4.2 \frac{T_{i+1} - T_{i-1}}{\Delta t}, \tag{4.44 a}$$

where G is the average daily soil heat flux in MJ m^{-2} d^{-1}, T equals the mean air temperature in °C for time period i, and t equals the time in days between the midpoints of the two periods. In real time predictions, where T_{i+1} is unknown, G for monthly or longer time periods can be predicted as:

$$G = 4.2 \frac{T_i - T_{i-1}}{\Delta t}. \tag{4.44b}$$

Additional information can be found in articles by van Wijk and de Vries (1963), de Vries (1963), and Brutsaert (1982).

For application to short-periods, e.g., when hourly data for G are required, other approaches must be used. One method commonly used in research is the use of heat flux plates installed near the soil surface (e.g., at 0.01 m). For full cover vegetation situations where temperature gradients even near the surface are minor, heat flux plates can give reasonably accurate data for G; however, for many situations, strong temperature gradients exist near the soil surface and differences in thermal conductivities of the plate and of the soil can produce large errors in estimating G. Also, with the strong thermal gradients during day and night, condensation and evaporation near the upper or lower surface of the flux plate can be a significant problem. Therefore, soil heat flux plates should be placed below depths of soil moisture evaporation (generally 0.08 to 0.15 m) to avoid errors associated with H to λE conversion between the surface and the plate.

In measurements of G in water, care must be taken to avoid effects of absorption of short-wave radiation by the plate. Therefore, the plate should generally be buried in the lake substrate. In applications where soil is saturated or is very dry, the user should check with the manufacturer to ascertain whether the plate calibration holds.

Measurements by soil heat flux plates, G_{zs}, at some depth z_s are usually adjusted to represent G at the soil or water surface by incorporating Equation 4.42:

$$G = G_{z_s} + C_s \int_0^{z_s} \frac{\partial T_s}{\partial t}\, dz, \tag{4.45}$$

where $\partial T_s/\partial t$ is the change in temperature within the z_s thick layer above the heat flux plate. Temperature is usually measured with thermocouples or buriable thermistors. Tanner (1960) developed resistance thermometry systems for obtaining the average 0 to 0.1-m soil temperature using a string of ten or more units to provide average temperature changes with time in the upper 0.1 m of soil.

Another method more widely applicable involves calorimetry to determine changes in soil heat storage within a zone from the surface to some depth z, where the flux is negligible or can be estimated reliably, e.g., with Equation 4.42. Both profiles of soil temperature and soil moisture content must be known making this method a difficult one at best albeit one of the most reliable, presuming reliable profile data are available.

Soil heat flux density for dense, short vegetation such as for grass, wheat, or alfalfa can often be approximated during daylight periods as $G = 0.1\ R_n$ (Clothier et al., 1986; Choudhury et al., 1987; Allen et al., 1994b). A universal relationship proposed by Choudhury et al. (1987) and Choudhury (1989) for estimating G under daytime conditions is:

$$G = 0.4\ e^{-0.5\,LAI}\ R_n \tag{4.46}$$

where LAI is the leaf area index and e is the natural number, with G having the same units as R_n. Equation 4.46 estimates $G = 0.1\ R_n$ for an $LAI = 2.8$, which is typical for clipped grass. Pruitt (1994), in an analysis of late spring and summertime data collected over clipped grass in Davis, California, found Equation 4.46 to provide close estimates of G during daytime for a range of LAI from around 2.8 up to 4.5. During moderate- to strong-advection days at Davis, G was found to be almost negligible or even reversed in direction during daylight hours. A very low fraction of G/R_n (0.036), was used by Verma et al. (1986) during daylight for a deciduous forest.

For nighttime under grass forage, Allen et al. (1994b) found $G = 0.5\ R_n$ to be valid. In Davis, a $G = 0.3\ R_n$ was found at night. In all of these reported ratios, it should be noted that the effects of seasonal variations are not addressed. The seasonal lag in soil or water bodies behind solar radiation, can be expected to produce higher ratios of G/R_n in spring than in fall months with equivalent solar or net radiation amounts.

In computerized evapotranspiration models where soil surface temperatures are calculated from surface heat exchange equations, G can be estimated using the "force-restore" method (Lin 1980). A thin surface layer is used to simulate surface temperature for use in estimating long-wave radiation and in "forcing" turbulent heat transfer to the air. A deeper second layer is used to "restore" the surface layer temperatures.

III. INTERACTION OF SURFACES AND METEOROLOGICAL FACTORS

A. Energy Balance—Air-Mass Interactions

1. Introduction Penman (1956) defined Potential Transpiration (evapotranspiration) as "the amount of water transpired in unit time by a short green crop, completely shading the ground, of uniform height

and never short of water;" however, others (Monteith, 1981) have defined potential transpiration or potential *ET* as evapotranspiration from vegetation having a saturated surface (surface resistance, r_s, = 0). The latter definition, although correct, has proven to be difficult to measure and to apply in the field, especially in the absence of weather data that reflect a potential *ET* environment (measured over a wet surface). Therefore, use of the term potential *ET* has been discouraged as a climatic index and for use in *ET* estimation.

Wright and Jensen (1972) indicated that maximum *ET* for nonsaturated conditions may be approximated by *ET* from a well watered crop of alfalfa with a height of at least 0.2 m. They suggested the use of the term reference crop *ET*. Doorenbos and Pruitt (1975; 1977) described a reference crop evapotranspiration (ET_o) as "the rate of evapotranspiration from an extensive surface of 8–15 cm tall, green grass cover of uniform height, actively growing, completely shading the ground and not short of water." Subsequently, this definition specified the grass cover as being that of one of the cool-season grasses with roughness, density, leaf area, and canopy resistance characteristics similar to perennial ryegrass (*Lolium perenne*) or alta fescue (*Festuca arundinacea* Schreb. 'Alta'), since warm-season grass varieties, such as Bermuda (*Cynodon dactylon*), exercise considerable control over transpiration.

Grass has served as a standard reference crop largely due to the preponderance of grass in lysimeter studies around the world. Also, in much of the world, several species of cool-season grasses remain green throughout the year. With grass, it is possible to maintain a near-constant canopy with a Leaf Area Index (*LAI*) of ~3, and with weekly or more frequent mowing, growth differences within and outside of a lysimeter, whether due to fertility or soil moisture differences, can be minimized.

Despite its widespread use, there are disadvantages in using grass as a reference. Because grass is short and aerodynamically quite smooth, the *ET* losses can be considerably lower than those for most other types of vegetation with rougher canopies, particularly in hot, dry, and windy conditions. Hence, well-irrigated alfalfa, some 0.3 to 0.5 m in height, has been the other major crop used as a reference with the symbol ET_r (Jensen, 1968; Jensen and Haise, 1963; Jensen et al., 1990). There are few vegetative canopies that may have higher *ET* rates than 0.3 to 0.5-m tall alfalfa. Thus, in a two-step procedure (using ET_r and a crop coefficient K_c), there should be, for full-cover conditions, a narrower range in K_c values than when ET_o (grass) is used as a reference. This would especially be the case when a wide range of climates and crops is involved.

There are also some disadvantages in the use of alfalfa as a reference. Cuttings destroy the entire canopy and actual evapotranspiration is drastically reduced for about a week following each cutting. Trampling of the crop can present much more of a problem than with a grass turf. Alfalfa can be difficult to grow in some tropical climates and at high altitudes if validation is required. In addition, various alfalfa varieties may exhibit different stem erectness and stomatal control. These variations may create differences in ET_r as large as 10% among alfalfa varieties (Wright, 1988).

The Working Group on Water Requirements of the International Commission on Irrigation and Drainage (ICID, 1985) provides an updated terminology listing for evapotranspiration and water requirements. They define reference crop evapotranspiration as: "evapotranspiration from a given well-adapted crop selected for comparative purposes under given weather conditions and with adequate fetch (sufficient to make edge effects relatively unimportant) and for a standardized watering regime appropriate for this crop and the region concerned."

The Working Group's definition of potential evaporation is: "the evaporation from a given surface when all surface-atmospheric interfaces are wet (saturated) so that there is no restriction due to either biological control or soil water content on the water vapor loss from the surface area." (ICID, 1985). They pointed out that its magnitude depends primarily on atmospheric conditions and surface albedo but also varies with geometric characteristics of the surface. In terms of hydrology, this term would apply only following precipitation, condensation (dew), fog interception, or a crop irrigated by sprinkling. Within Chapter 4, potential evaporation is identified as E_o and relates to the above definition.

In the following section, data are presented and equations are used to illustrate, in general, some of the basic interactions relative to *ET* that take place between plants and the atmosphere.

2. Latitude, Temperature, Humidity and Wind

a. R_n vs R_s Relationships. Net radiation of cropped surfaces during daylight is primarily a function of incident solar radiation. A number of the important *ET* equations involve the input of total 24-hour net radiation, which is related not just to R_s but to the length of the night and to nighttime surface and atmospheric conditions. The examples in Fig. 4.1 for grass or grass-clover crops show that there can be a wide range in the fraction of solar radiation involved in 24-hour net radiation, although it is evident from Fig. 4.1(a) that in spite of the latitude differences and obvious differences in climate that must exist between Copenhagen, Davis, California, and Zaire, the data could be represented by a single regression relationship. The offset indicated relates to the need for some 3 to 5 MJ m^{-2} day^{-1} of R_s at Copenhagen and Davis, respectively, to make up for the negative R_n of the long nighttime periods of winter. Even though r^2 values exceed 0.92, the results presented in Fig. 4.1(b) show a wide range of R_n/R_s fractions during the year. The Copenhagen results are unique in that the long winter nights can result in negative R_n exceeding in magnitude the daytime positive R_n, thus producing a negative fraction of R_n/R_s. The 10-year results of Aslyng and Jensen (1965) reveal an average negative 24-hour R_n for the four months of November through February.

The only location where R_n/R_s remains fairly steady throughout the year is at the tropical site (Zaire). The slightly lower midsummer fractions in Zaire as compared to the other locations probably relates to the 12-hour days in Zaire with much longer days at the higher latitudes.

Because of the wide variation of R_n/R_s with seasons except in the tropics, a requirement of considerable empirical adjustment might be expected in the use of solar radiation-based evaporation equations; however, in interpreting the R_n/R_s patterns of Fig. 4.1, it should be noted that except for moderate to strong wind conditions, *ET* is largely a daytime phenomena, and hence can be expected to be more strongly related to daytime net radiation than to 24-hour total R_n. The $R_n^{\dagger}/R_s$ curve for Davis in Fig. 4.1 (b) reveals rather minor seasonal variability, especially in comparison to R_n/R_s for Davis. Thus, a closer correlation between *ET* and R_s can be expected than that which might be inferred from the R_n/R_s relationships.

b. Effect of Weather on Fraction of R_n used in ET. A-priori reasoning might lead one to conclude that for an effectively wet surface (i.e., a very thin film of water on the surface or a complete canopy of vegetation with insignificant stomatal resistance), most of the net radiation minus soil heat flux ($R_n - G$) would be converted to latent heat λE with only minor amounts of sensible heat, *H*. Surface temperature should remain at or near air temperature, and near-neutral stability conditions within the turbulent boundary layer might be expected during much of the daytime period. Just such a concept concerning the partition of net radiation for vegetative surfaces was prevalent in earlier literature. Researchers suggested potential evapotranspiration [as defined by Penman (1948) or by Thornthwaite (1948)] as a condition under which essentially all of the ($R_n - G$) would be used in the evaporative process; however, in reality this is rarely the case, although the effect of advection in producing ratios of latent heat, λE to R_n well over 1.0, has long been recognized. In the early literature, ratios less than 1.0 seemed to be expected only where significant stomatal resistances were likely, or where plant cover was only partially shading the ground or just upwind from cold ocean or lake surfaces.

The following examples of seasonal patterns of *ET* and of $\lambda ET/R_n$ reveal that the above assumptions were far too simplistic. The results shown in Fig. 4.2 are for studies where, for the most part, full cover, vegetative surfaces with low stomatal resistances were involved. Data for September and October in Phoenix, Arizona are missing due to lack of full cover and/or low soil moisture conditions. For Davis, 100 out of over 1200 days on record were excluded from the analysis because of strong advection. The Copenhagen curve is based on 1970 daily data (Jensen, 1972) as supplied to FAO for development of Irrigation and Drainage Paper 24 (Doorenbos and Pruitt, 1975; 1977).

The climates of Phoenix, Davis, and Copenhagen, due to latitude and other factors, represent a wide range of weather conditions. Although advection (both heating and cooling) was no doubt a factor at times, the normal seasonal lag of temperature and saturation vapor pressure deficit behind radiation

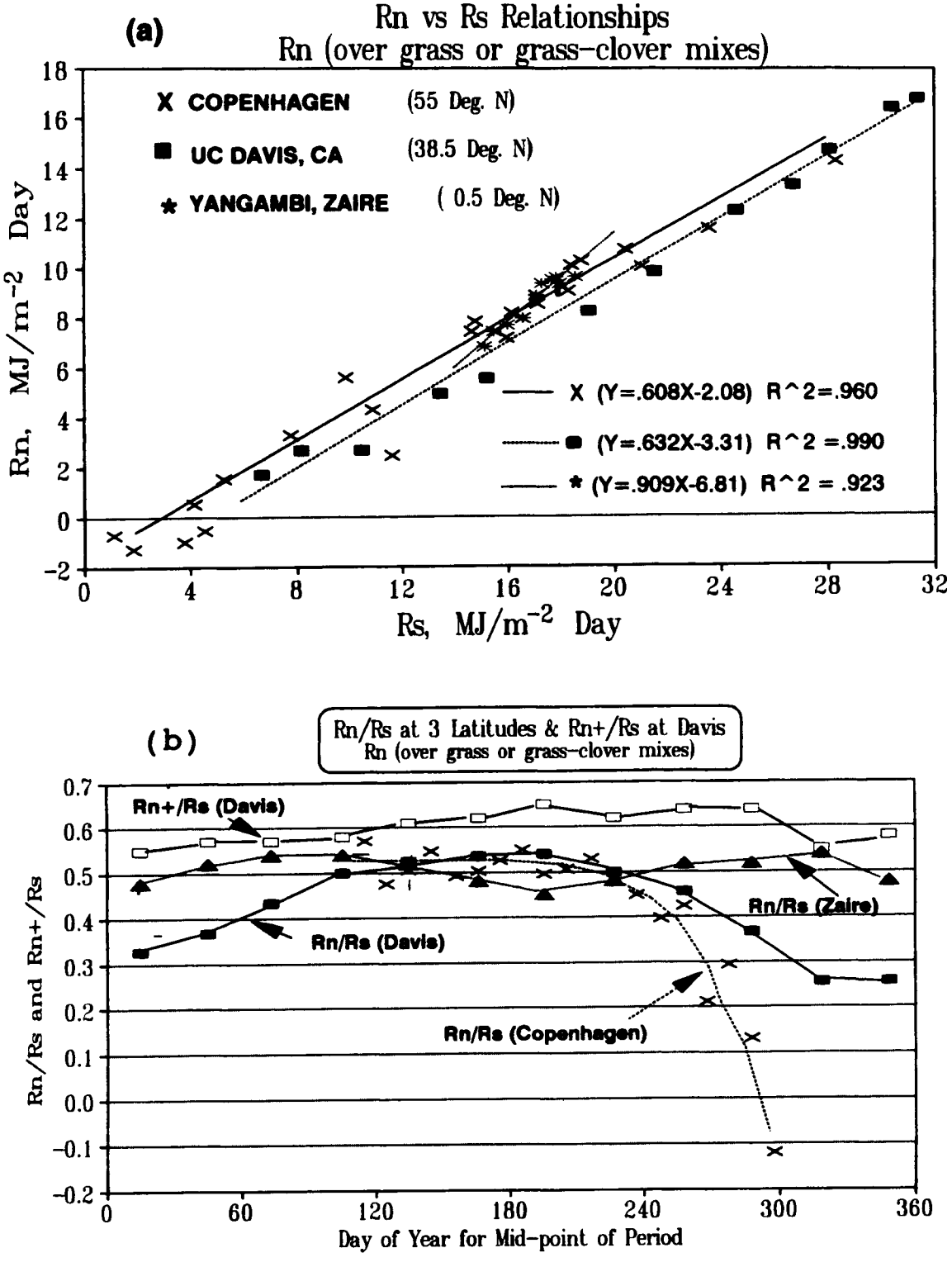


Figure 4.1—(a) Regression Relationship of R_n vs R_s for Three Locations; (b) R_n/R_s and R_n^+/R_s Fractions as a Function of Day of Year for Copenhagen, Denmark, Davis, CA, and Yangambi, Zaire (Jensen, 1972; Pruitt, 1971; Bultot and Griffiths, 1972).

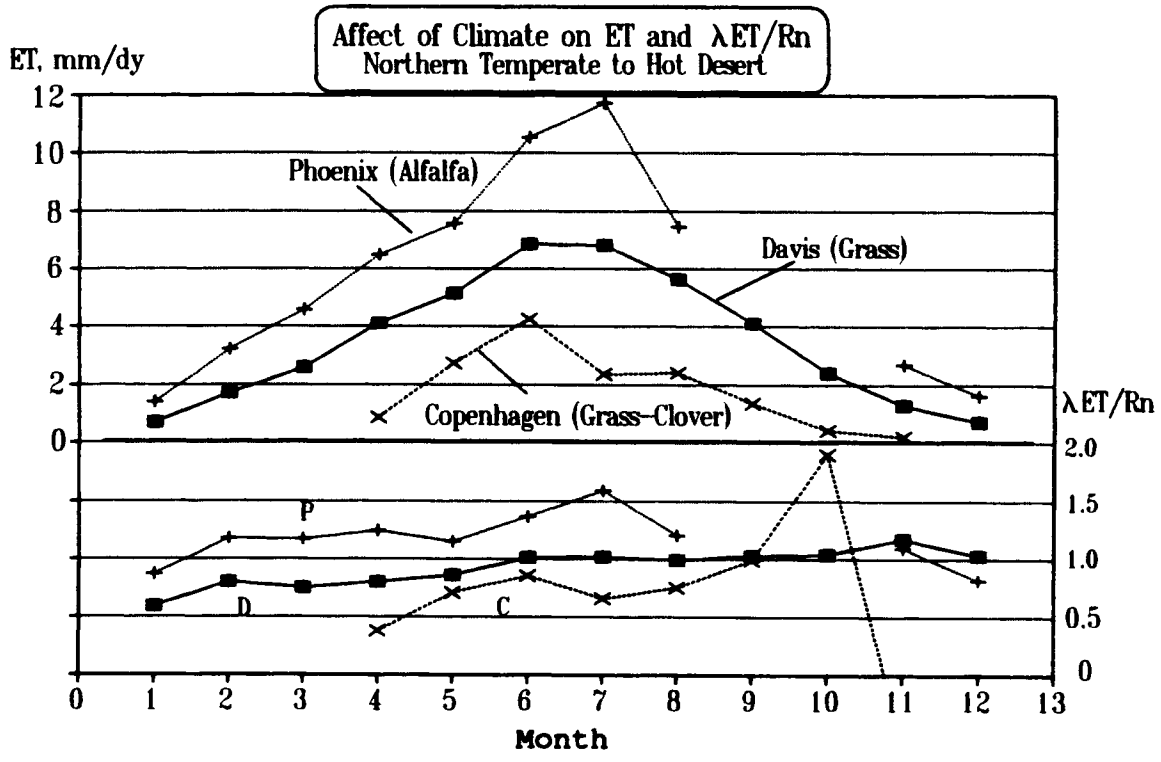


Figure 4.2—Evapotranspiration and the Fraction of 24-hour R_n Involved in λET for Cropped Surfaces at Phoenix, AZ, Copenhagen, and Davis, CA (Pruitt, 1971; Van Bavel, 1966).

produced a strong trend for lower $|\lambda E/R_n|$ in late winter and spring months than for summer and fall months.

In Fig. 4.2, curves for $\lambda ET/R_n$ were based on 24-hour values of R_n. In Fig. 4.3(a), Davis, California data for January 1960 to June 1963 show not only the fractions $\lambda ET/R_s$ and $\lambda ET/R_n$, but the fraction of daytime positive net radiation used in *ET*, $\lambda ET/R_n$. During the last half of the year, the lower curves are distinctly different in shape than the upper curve based on total 24-hour net radiation. Since days of strong advection were not included in this analysis, the data are primarily for days when some 95-100% of the total *ET* occurred during daylight hours. The $\lambda ET/R_n$ curve thus reveals that well-irrigated, frequently-mowed grass in the climate at Davis does not, in the daytime, use up all the net radiation as latent heat, and that significant sensible heat is lost to the air and soil during daytime periods, especially during non-summer months. This has important implications, not only for the energy balance, but also in relation to stability conditions that prevail during daytime within the turbulent boundary layer. This in turn relates to the question of whether or not stability-related correction expressions are required for accuracy in estimating *E* or *ET*.

The results in Fig. 4.3(b–d) show a clear seasonal cyclic or lag effect in *ET*-radiation relationships. One of the early papers revealing such trends (Pruitt, 1964) was based on only 1960 out of the 3-½ years of data presented in Fig. 4.3.

From the preceding figures it is obvious that the net radiation partitioned into latent and sensible heat fluxes is highly dependent upon prevailing air-mass conditions. Presenting a basic understanding of the effect of various climatic factors on *ET* is difficult. We would like simple answers to questions such as "what is the effect of wind on evapotranspiration, or of radiation, or of humidity?" However, it is obvious that these and other factors, including surface conditions and the level of radiation itself, are all interactive.

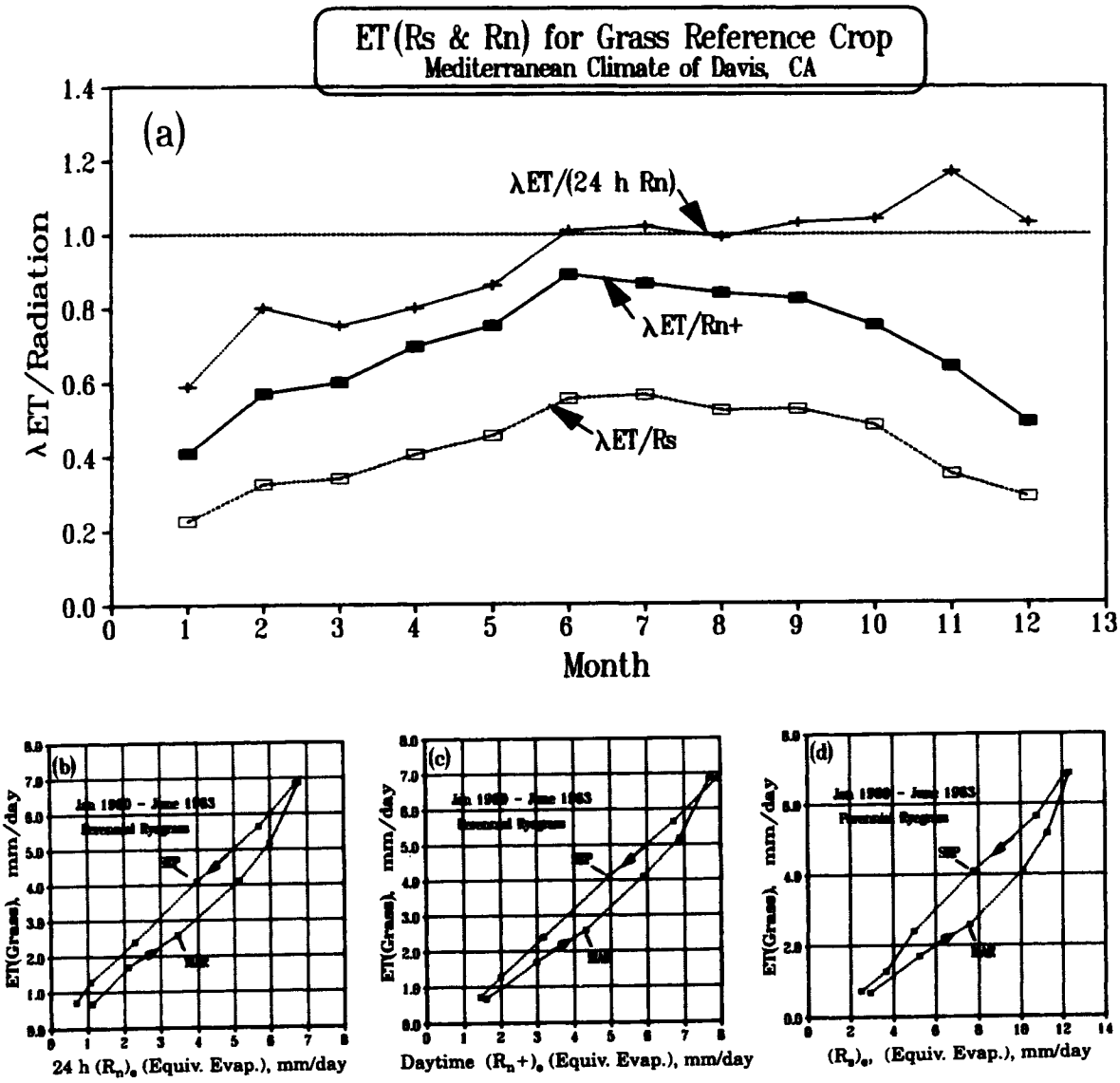


Figure 4.3—(a) Mean Monthly Fractions of Solar and Net Radiation Used as λET by Perennial Ryegrass and (b – d), the Cyclic relationships of ET Versus $(R_n)_e$, $(R_n+)_e$, and $(R_s)_e$, (equiv. mm. evap.), Jan 1960 - June 1960-63. Davis, CA (Pruitt and Angus, 1960; Pruitt, 1971).

3. ***Equations Illustrating Weather Effects on Potential Evaporation*** The combination energy balance-aerodynamic equations, such as those proposed by Penman (1948) or Slatyer and McIlroy (1961), offer excellent models for illustrating how the various weather parameters interact to affect energy partitioning at evaporating surfaces. These models are based on empirically-determined aerodynamic wind functions while the more theoretically based Penman-Monteith (Monteith, 1965) method has a wind function term based on the log-law wind profile, and includes a leaf or canopy resistance term.

The Penman-Monteith equation will be used later to present examples which help in evaluating not only the effects of air-mass conditions, but surface resistance as well, on net radiation partitioning at the surface into the various energy balance terms. At this point, the simpler McIlroy equation is used to cover some aspects of the above and to introduce a now widely-used method for estimating *ET* for non-advection conditions (Priestley and Taylor, 1972).

Since the McIlroy equation is used merely to illustrate the weather effects, units for the terms are not needed. It can be expressed as:

$$\lambda E = \left[\frac{s}{s+\gamma}\right](R_n - G) + h\,(D_z - D_s), \tag{4.47}$$

where λE is the latent heat flux density, γ is the psychrometric constant, and s is the slope of the saturation vapor pressure versus temperature taken at the average of surface temperature and the wet-bulb temperature of the air at height z. D_z is the wet-bulb depression at height z, D_s is the same for air at the surface, and h is a wind-speed dependent heat-transfer coefficient which might be expected to vary with stomatal resistance, crop roughness, and stability conditions of the air mass above the surface. For a hypothetical effectively moist surface, D_s would drop out and, except for the empiricism of the wind-function term normally determined under non-wet surface conditions, the equation would be a predictor of potential evaporation, E_o, as defined earlier (ICID, 1985).

Considering the effects of temperature alone on the energy balance, the term $s/(s+\gamma)$ in Equation 4.47 is temperature sensitive, e.g., at a barometric pressure of 101.3 kPa, varying from 0.4 at 0°C to 0.85 at 40°C. Hence, under very humid conditions (or climates) where the second half of the equation would have little significance, the equation predicts that the fraction of R_n used as λE could be expected to be greater in the warm fall months as compared to cooler spring months, or for cold or cool morning periods as compared to afternoon periods. Also in less humid climates the warmer fall temperatures (or afternoon periods) would result in greater wet-bulb depressions, D_z, producing even higher fall or afternoon ratios of $\lambda E/R_n$. In locations well away from the equator, well-defined cyclic effects can be expected in the relationships of λE vs. R_n (or R_s) on an annual basis (Fig. 4.3) as illustrated for both annual and diurnal periods more than three decades ago by Pruitt (1964).

An interesting case to consider, and one which can provide an excellent estimate of the lowest $\lambda E/R_n$ to be expected for potential evaporation cases, is the use of the first half of Equation 4.47. This portion of the equation offers an estimate of so-called "Equilibrium Evaporation," a condition described by Slatyer and McIlroy (1961) for cases where air in the lower meter or so is saturated or nearly saturated. Predicted effects of wind become zero, (except as it affects temperature) and the ratio of $\lambda E/R_n$ is defined by $[s/(s+\gamma)]$, a function of temperature alone for any given barometric pressure. Such conditions would seldom exist for a daily or longer period; however, on a diurnal basis in many climates, *RH* is nearly 100 percent for the first hour or so of sunlight.

The well-known Priestley-Taylor (1972) equation essentially provides a compromise between "equilibrium evaporation" and the requirement of including an aerodynamic term such as that in Equation 4.47. The Priestley-Taylor equation can be expressed as:

$$\lambda E = \alpha\left(\frac{\Delta}{\Delta+\gamma}\right)(R_n - G), \tag{4.48}$$

where Δ is the slope of the saturation vapor pressure vs air temperature relationship and α is a multiplier which essentially compensates for the lack of an aerodynamic term typical of combination aerodynamic-energy balance equations. Note that the $s/(s+\gamma)$ of McIlroy's equation is equivalent to the more commonly used $\Delta/(\Delta+\gamma)$, although the temperature base used for the latter is the dry-bulb air temperature.

Priestley and Taylor analyzed data obtained in earlier studies conducted over both land and ocean surfaces and found that a value for α of 1.26 in Equation 4.48 provided estimates of *ET* for the land studies in good agreement with measured *ET*; however, this was the case only when data from land-based studies were restricted to include the first few days immediately following rain over an extensive upwind area. See Jensen et al. (1990) for a comprehensive analysis of this and other methods using world-wide data.

4. Summary Although net radiation at an evaporating surface is no doubt the most important single factor affecting *ET* over daily or longer periods, studies in the early 1960s clearly showed a strong out-of-phase relationship between *ET* and R_n, both diurnally and seasonally. It is clear from the results considered herein that the partitioning of net radiation into various components of the energy balance is very much a function of other climatic factors such as wind, temperature, and humidity as well as plant and soil factors. Actual *ET* by crops can be significantly affected by stomatal resistance as illustrated by the Penman-Monteith equation, where under some conditions increased wind may even result in decreased *ET*. The degree of plant cover, stage of maturity, and moisture status of the soil are also important factors.

IV. EVAPORATION FROM WATER SURFACES

A. Introduction

Evaporation from water surfaces is a major component of the hydrologic cycle and must be considered in a wide variety of hydrologic and water resource studies. Because of its nature, evaporation from water surfaces is rarely measured directly, except over relatively small spatial and temporal scales (Jones, 1992). Evaporation from a water surface is most commonly computed indirectly by one or more techniques. The techniques described here are applicable to most engineering and hydrologic studies and include measurements from evaporation pans, water balance, energy balance, mass transfer procedures, and a combined technique. The selection of the best technique to use for a particular computation is largely a function of the data availability and the required accuracy of the computed evaporation.

B. Methods of Estimating Water Surface Evaporation

1. Pan Evaporation Evaporation from water bodies can be estimated from a network of evaporation pans sited throughout the country. The standard evaporation pan in the United States is the Class A Pan which is 1.21 m in diameter and 0.254 m deep placed 0.15 m above ground level on an open-timber framework. Two other commonly used pans are the Soviet GGI-3000 (0.3 m^2) pan and the GGI-20 m^2 tank, both placed in the soil with only 0.075 to 0.1 m of rim above the soil surface. Details can be found in the "Guide to hydrometeorological practices" (World Meteorological Organization 1970). Daily evaporation is computed from the water balance of the pan, which is the difference between the change in storage in the pan and inputs of precipitation and water added to maintain an adequate supply in the pan for evaporation. Due to differing thermal characteristics between the pan and large water bodies, the pans tend to overestimate the total amount of evaporation and distort the seasonal distribution. Thus, on a seasonal basis, pan evaporation usually peaks several months before the peak evaporation of deep lakes. Pan estimates are also suspect during late fall and early winter periods when ice may form in the pan but not on large water bodies. Annual lake evaporation estimates are obtained by multiplying the annual pan data by an appropriate coefficient. These coefficients have been computed for a number of water bodies and, for a United States National Weather Service (NWS) Class A pan, tend to range in value from 0.65–0.85 (World Meteorological Organization, 1973; U.S. Department of Commerce, 1968). The coefficient is higher under humid conditions and lower under arid or dry conditions. A coefficient of 0.70 is applicable when water and air temperatures are approximately equal; however, research over the last several decades points to the need for caution in the use of general pan evaporation data to estimate lake evaporation or crop evapotranspiration (*ET*).

Early studies in the United States in semi-arid to arid climates revealed a very significant response in evaporation to pan size, but now it is clear that this result was primarily due to a decreasing effect of local advection with increasing size of the water surface area. Pruitt and Doorenbos (1977b) reported almost no difference (4 percent) between evaporation from a 0.62-m diameter Soviet pan and a 5-m diameter Soviet pan when both were located in a 5-ha irrigated grass field. On the other hand, at a dry site in

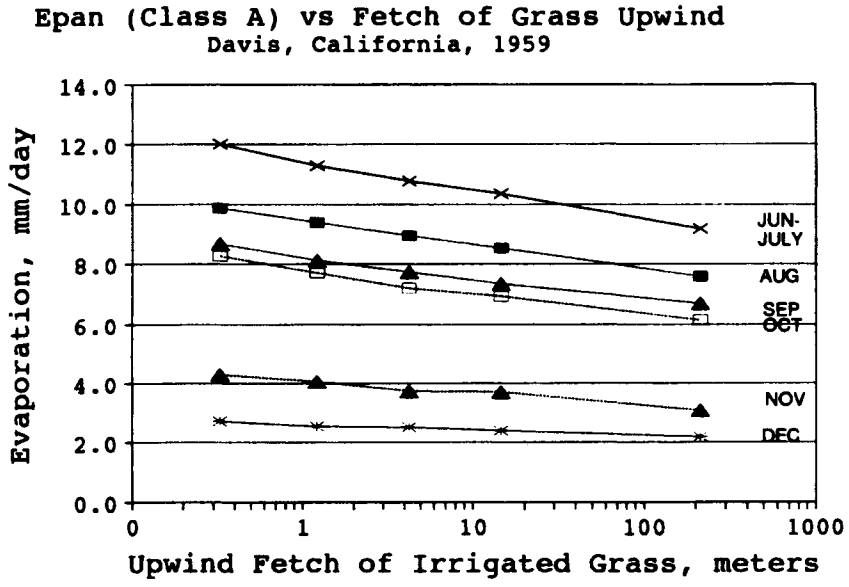


Figure 4.4—Evaporation from NWS Class A Pans Plotted Against the Upwind Fetch of Irrigated Land Cropped to 0.07 to 0.15-m tall Grass, Davis, CA (Pruitt, 1966).

Nevada, evaporation from the smaller Soviet pan averaged some 1.6 times that of the larger pan, although when corrected for the net heat transfer from soil to the pans, a factor of 1.45 resulted (Hounam 1973).

Young (1947) presented a good discussion of the problem of local pan environment in relation to estimating lake evaporation. Soon thereafter, studies in India by Ramdas (1957) and studies at Prosser, Washington and in California (Pruitt, 1960; 1966; State of California, 1975; 1979) provided clear evidence that unless local environment of pans was taken into account, the estimation of crop *ET* (or of lake evaporation) was subject to errors of up to 35%. Fig. 4.4 presents some of the early results that illustrate the problem associated with the environment just upwind. These results came from a study involving four Class A pans located within a large fallow field, with three of them placed within various-sized circular areas, flood irrigated frequently and planted to grass. The fifth pan located within a 5-ha irrigated grass field had a minimum upwind fetch of grass or irrigated pasture land of some 200 m. Data from the one pan in the fallow field that had no surrounding grass were plotted as if having a 0.3-m fetch of grass in order to use a log-linear plot. From this and similar studies (e.g., Ramdas, 1957; Pruitt, 1960; Stanhill, 1962), Pruitt developed recommended pan K_p values for estimating ET_o and in turn evaporation from shallow water bodies. Taken into account were the effects on pan coefficients of upwind fetch (both dry and moist), mean relative humidity, and total daily wind on K_p (Jensen, 1974; Doorenbos and Pruitt, 1975; 1977; Jensen et al., 1990) (see Table 4.10 later in this chapter).

Mean monthly, seasonal, and annual Class A pan evaporation for individual stations throughout the United States have been tabulated by Farnsworth and Thompson (1982) and Farnsworth et al. (1982). The second reference, "Evaporation Atlas for the Contiguous 48 States," contains maps with isolines of pan coefficients recommended for use with the evaporation maps in estimating so-called "free water surface evaporation" (FWS) from shallow water bodies. Fig. 4.5 presents a portion of the Annual FWS Evaporation map for the 48 states from that publication. Seasonal maps for the periods May–October and November–April are included in this National Weather Service report.

2. ***Water Budget Procedure*** The water budget procedure is a simple method for computing evaporation from a water surface for monthly, seasonal, or annual time periods. All of the terms in the water

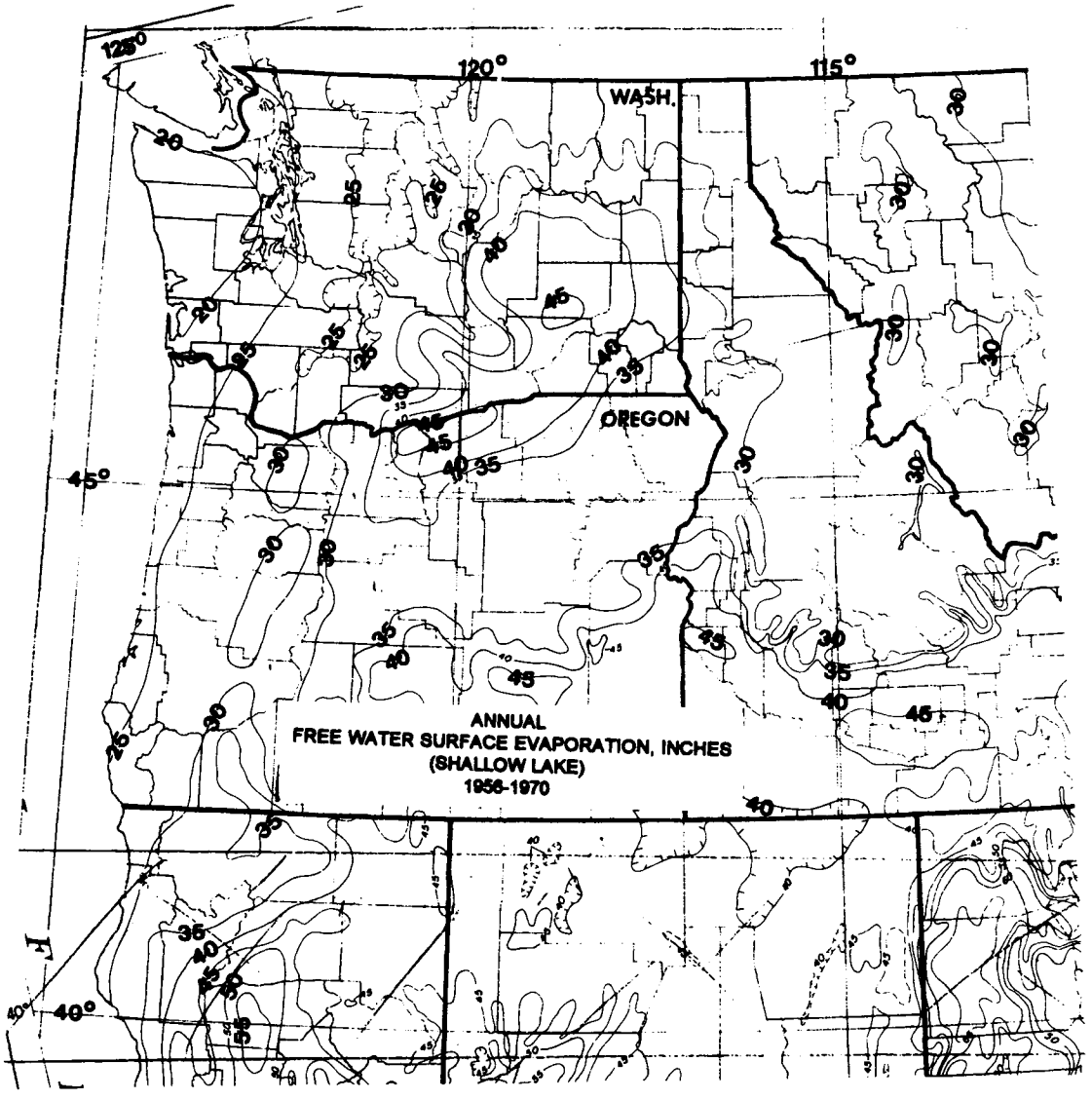


Figure 4.5—A portion (NW USA) of a Map of Estimated Free Water Surface (FWS) Annual Evaporation for the 48 States (Farnsworth et al., 1982).

balance, except for evaporation, are either measured or estimated with evaporation computed as a residual. The water balance procedure is based upon the hydrologic equation

$$Inflow = Outflow + \Delta Storage. \tag{4.49}$$

Inflow terms consist of precipitation on the water surface, runoff from the land basin, major channel inflow from outside the immediate drainage basin, ground water inflow, and diversions into a free water body (lake, reservoir, pond, etc.) from outside its basin. Outflow terms consist of evaporation from the water surface, major channel flow out of the water body, diversions out of the water body, and ground water flow from the water body. Thus, the water balance equation can be expressed in terms of water depth evaporated from the surface, E, as:

$$E = P_r + R + Q_i + G_w + D_i - Q_o - D_o - \Delta S, \tag{4.50}$$

where P_r is the precipitation on the water surface, R is the runoff inflow from the drainage basin, Q_i and Q_o are major channel flows into and out of the water body, respectively, G_w is the net ground water flow into the water body, D_i and D_o are diversions into and out of the water body, respectively, and ΔS is the change in water body storage.

For most calculations, the terms are expressed in units of depth in mm or cm over the water body surface for a specified time interval. The computations are most accurate when the evaporation is of the same order of magnitude or larger than the other terms of the water balance. The ground water component is usually the most difficult term to estimate.

Water balance data sources are the National Climatic Data Center, NOAA regional data centers and state climatologists for precipitation data, the U.S. Geological Survey, state and provincial water agencies, and flood and conservation districts for runoff, major channel flows, and diversion data.

3. Aerodynamic Method Aerodynamic methods are among the most widely applied to determine evaporation from lakes and large reservoirs. The basic equations were tested on Lakes Hefner and Mead in the 1950's (U.S. Geological Survey, 1954; 1958; Harbeck, 1962). The basic Lake Hefner form is

$$E = M\,(e_s - e_z)\,u_z, \tag{4.51}$$

where E is the evaporation in mm per day, M is the mass transfer coefficient, e_s is the saturation vapor pressure at the surface water temperature T_s (kPa), e_z is the vapor pressure of the air at level z (kPa), and u_z is the wind velocity at level z in m s^{-1}. The vapor pressure of the air e_z is equivalent to the saturation vapor pressure of the air at the dew point temperature and is usually computed by

$$e_z = 0.01\,RH\,(e_z^o), \tag{4.52}$$

where e_z^o is saturation vapor pressure at air temperature (T_z) at level z (kPa), and RH is the relative humidity (%). Values of e_z^o can be obtained from standard meteorological tables (List, 1963; Berry et al., 1973), or estimated from Equation 4.1 using water surface temperature for T_s.

The mass transfer coefficient, M (Equation 4.51), is a function of the measurement height of the meteorological measurements, the atmospheric stability, barometric pressure, roughness of the water surface, size of lake, and the density and viscosity of the air. For the Lake Hefner study (U.S. Geological Survey, 1954), M, an empirical coefficient, is 0.097 for wind, temperature, and humidity measurements at the 8 meter level and where vapor pressure is expressed in kPa. The basic mass transfer equation requires measurements of water temperature, air temperature, humidity or vapor pressure, and wind speed. The wind speed and vapor pressure gradients can be adjusted from one level to another level over the same surface by the logarithmic law (Oke, 1978; Allen et al., 1989; Jensen et al., 1990), provided that the effect of stability is negligible:

$$u_2 = u_1 \frac{\ln\left(\frac{z_2 - d}{z_{om}}\right)}{\ln\left(\frac{z_1 - d}{z_{om}}\right)}, \tag{4.53}$$

where z_1 and z_2 are the measurement heights for level 1 and level 2 (m), respectively, z_{om} is the roughness height for momentum transfer (m), and d is the zero-plane displacement (m). For many water surface applications, z_{om} can be taken as 0.0001 m and d is generally 0 for water. The wind level adjustments are relatively insensitive to the value of z_{om}. The difference between using 0.0001 m and 0.001 m is approximately 5 percent which is well within the application accuracy.

Most mass transfer research over the past 25 years has been directed towards the bulk aerodynamic equation. The bulk equations can consider the impact of atmospheric stability as well as air mass characteristics on the evaporation rate. This form of equation can be written as

Chapter 5

GROUND WATER

I. INTRODUCTION

Ground water comprises the subsurface water that occurs within the saturated zone of the earth's surface, or within the part of the earth's crust where all openings are filled with water. It is distinguished from water on and above the earth's crust, and from water occurring below the ground surface but above the saturated zone (or occurring within the unsaturated zone). Nevertheless, ground water is only one phase of the hydrologic cycle and cannot be considered separately or in isolation from all the other phases, especially when managing water resources of an area.

The amount of water stored in the earth's crust is immense. The volume of water stored in the rocks of the world's land areas may be on the order of 8 billion km^3, of which half is at depths less than 800 m. This volume is about 35 times the combined storage of all the world's rivers, fresh water lakes, reservoirs, and inland seas, and is about one-third the volume of water stored in arctic and antarctic ice fields, in the glaciers of Greenland and great mountain systems of the world.

Despite its magnitude, most ground water is not visible because it is stored below the earth's surface; however, a small percentage of this underground supply is visible as naturally-discharging springs, or as streams and rivers which are sustained by ground water when direct runoff from precipitation ceases.

This lack of visibility tends to promote some mysticism about ground water and a disregard for the need to protect it from pollution. Ground water is not mystic, and it does need protection. While ground water is a resource of immense magnitude, it is not evenly available and is exhaustible. Excessive continued exploitation may result in damage to ground water resources through lower ground water levels, aquifer consolidation and resulting land surface subsidence, water quality degradation, and plain exhaustion of the supply. Spilling or dumping of wastes on the earth's surface invites pollution of ground water. It is easier and much less costly to prevent pollution of ground water than to restore its quality after pollution.

This chapter presents the basic hydrologic facts about ground water and some of the management techniques that are used to maintain the viability of ground water resources. Most of this material has been taken directly from ASCE Manual 40 "Ground Water Management," 1987 edition. ASCE Manual 40 contains further discussion of ground water aspects related to design and operation of artificial recharge projects, surface and subsurface exploration and testing, development and use, legal and organizational considerations, use of aquifers for purposes other than water supply, land-surface subsidence, economics and implementation of management plans, and monitoring.

II. SOURCE AND OCCURRENCE OF GROUND WATER

On the basis of the sources from which they are derived, subsurface waters may be classified as meteoric, magmatic, connate, and metamorphic (White, 1957). Meteoric water is the water in the circulatory system of the hydrologic cycle, which includes evaporation from the water bodies and soil of the earth's surface, transport as vapor in the atmosphere, precipitation, overland runoff, stream flow, infiltration into the soil mantle, and downward percolation, as well as subsurface movement. Practically all the fresh water available comes from this circulating system, and meteoric waters are therefore of dominant interest to hydrologists.

For fresh water supplies, the remaining classes are of little practical value, although they are of considerable concern when encountered in wells as they generally are too saline for potability. Magmatic waters derive from molten rock (magma), and are found in some volcanic areas and some thermal springs. Connate waters were trapped in interstices of sedimentary rocks at the time of deposition, or at a later period that is still, in a geological sense, in the distant past. These waters may be encountered in deep wells (notably oil wells, where they commonly are found as brines) and in some mineralized springs. Metamorphic waters are those waters brought into existence by the heat, pressure, and recrystallization that created metamorphic rocks. They also may be encountered in some springs. These last three classes of water are generically related, respectively, to the three major classes of rocks (igneous, sedimentary, and metamorphic), and may similarly grade one into another. Some classifiers of subsurface water also list so-called juvenile waters, which have existed in the deep interior of the earth and are believed to be approaching the surface for the first time, but such waters have no characteristics that would distinguish them from the classes mentioned previously.

Subsurface space may be divided vertically into zones based upon the relative proportion of pore space that is occupied by water: a "saturated zone" in which all pores are filled with water, and an overlying "unsaturated zone" in which the pores contain gases (chiefly air and water vapor) in addition to any water that may be held by molecular attraction. During recharge periods, water moves under the force of gravity downward through the unsaturated zone to the saturated zone below (Remson et al., 1959). The water table has been defined as the upper limit of the saturated zone and is that locus of points in which the ground water is at atmospheric pressure. So far as water use is concerned, practically all vegetation, except for phreatophytes, depends on water in the unsaturated zone because the roots require aeration as well as water. In contrast, wells, springs, and streams (and therefore, people) obtain water from the saturated zone.

The upper part of the unsaturated zone, immediately beneath the land surface, first receives the infiltration from precipitation or some other source of surface water supply. This zone of soil water, extending downward from the land surface as far as plant roots penetrate, is of prime interest to agriculturists because of its ability to retain water for plant use. In places where the surficial materials are gravel, coarse sand, or very permeable rock, water retention may be negligible, and hence there is likely to be little or no vegetative cover present.

Beneath the zone of soil water is an intermediate zone, where water is retained solely by molecular forces, and where downward percolation presumably occurs, at least intermittently, toward the saturated zone. This zone comprises an indefinite zone in which there is no practical means of extraction. In some places, this zone is hundreds of meters thick; in others it is completely absent. In the latter case, the water table is found within the soil water zone.

At the base of the unsaturated zone, water may rise by capillary forces, the amount of rise ranging from practically zero in coarse material to as much as 2.5 m or more in clay materials. All pores may be saturated near the base of this capillary fringe; the number of pore spaces that are filled with water decreases in an upward direction. Water in the capillary fringe exists at pressures less than atmospheric pressure. In areas of shallow water table, the capillary fringe may extend upward to the root zone of plants and even to the land surface, thus permitting discharge of water by evapotranspiration. Because of the variation in capillarity among pores of different sizes, the upper limit of actual saturation must be

an irregular surface, difficult to observe or to delineate. Therefore, so far as practical use is concerned, it is desirable to refine the definition of a water table so that it marks the upper limit of saturation at atmospheric pressure. Thus, its position can be observed in supercapillary openings such as wells, sumps, shafts, and caves penetrating below the water table.

The saturated zone extends downward from the water table to depths where interconnected pores still exist. Such depths vary with compression of the rocks and appear to exceed depths penetrated by oil wells (more than 12,000 m). As a rule, the depth reached by the fresh circulating water included in the concept of the hydrologic cycle is much less, although some wells obtain fresh water at depths exceeding 1,000 m below the land surface.

Numerous complications are introduced into this simple picture by the great variations in the size of pores and in the rate of water movement through them. Because the permeability of rock materials varies so much, water may accumulate in some parts of the unsaturated zone sufficiently to cause saturation, at least temporarily, in a perched zone, which would then form a perched water table. Within the saturated zone, some rocks are of such a low permeability that gravity water movement is negligible and no determinable water table exists.

Geologic materials that are sufficiently permeable to transmit water readily under a pressure differential and which contain water are called aquifers. By contrast, aquitards are the materials that are sufficiently impermeable to preclude, or severely restrict, such movement for all practical purposes. An important result of the great variability in permeability of material is the occurrence of water in confined and unconfined states.

Among the saturated materials capable of yielding water to wells, springs or streams, a classic distinction is made between unconfined (or water table) aquifers and confined (or artesian) aquifers.

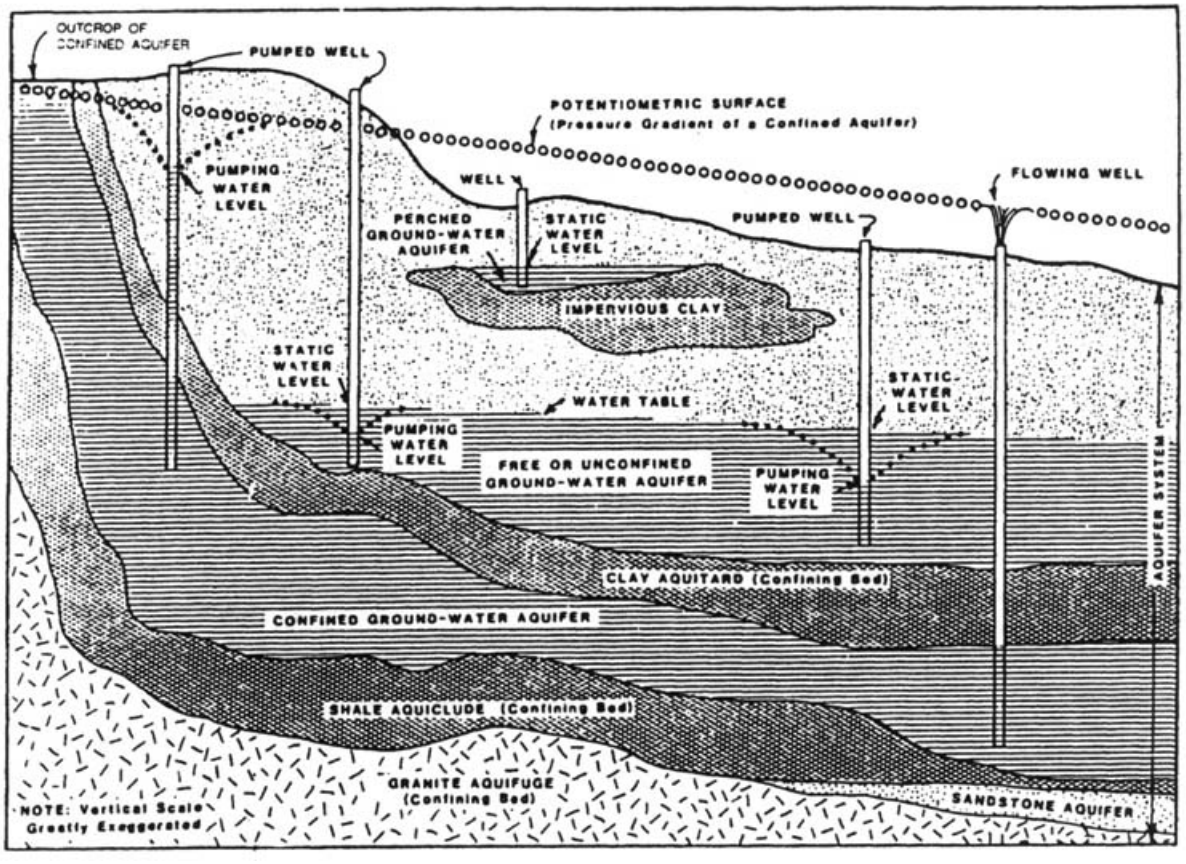


Figure 5.1—Diagrammatic Cross Section Showing Free, Confined, and Perched Ground Water Conditions

Unconfined ground water is found in the saturated zone whenever the upper surface of the zone forms a water table under atmospheric pressure, free to rise and fall with changes in volume of stored water. Confined water is found in aquifers which are separated from the unsaturated zone or another saturated zone by materials of markedly less permeability (see Fig. 5.1). The water pressure throughout the saturated thickness of a confined aquifer is everywhere in excess of atmospheric pressure.

Obviously, the distinction between confined and unconfined water results chiefly from the variations in permeability of geologic materials. Variations in permeability are commonplace. Although there are many exceptions, unconfined ground water is most likely to be found only in the shallowest wells, while water in deep aquifers is likely to have some degree of confinement. Because permeability is a relative term, confinement is also relative. Materials with identical characteristics may form an aquifer in one place and a confining layer above a much more permeable aquifer in another place.

An artesian well is one in which the water level in the well rises above the top of the aquifer in which it is contained. If the water rises above the land surface, a flowing artesian well is formed. Pumping of confined water causes rapid pressure reduction in the aquifer for considerable distances from the discharging well.

For hydrologists concerned with quantitative determination of ground water storage and movement, an important reason for distinguishing between confined and unconfined water is that the hydraulics of flow involved are different and require different testing techniques and formulas, as outlined by D.K. Todd (1980). The difference is comparable, in many respects, to that between flow in pipes and open-channel flow. It should be noted that the conditions of confinement can be changed. Pumping from wells in confined aquifers may eventually dewater the upper part of the aquifer, first in the immediate vicinity of the wells and then in progressively broader areas, until the entire aquifer becomes unconfined. The converse may also occur in recharge operations where water is injected into aquifers overlain by less permeable rocks.

For confined aquifers, the material is not dewatered or drained unless the hydraulic head drops below the top of the aquifer. Consequently, the term "specific yield" does not apply to confined conditions. A correlative term, the storage coefficient or storativity, is applied to confined aquifers and is defined as the volume of water an aquifer releases from or takes into storage per unit of surface area of the aquifer per unit change in head. Storage coefficients are usually on the order of 5×10^{-2} to 1×10^{-5}.

The water obtained from a confined aquifer results from a slight expansion of the water and a very small compression of the porous medium. Artesian aquifers are apparently all more or less compressible and elastic, though they differ widely in the degree and relative importance of these properties. In general, the properties of compressibility and elasticity are of most consequence in aquifers that have low permeability, slow recharge, and high head. If it were not for the elasticity, the supplies of water that could be recovered through artesian wells for intermittent or fluctuating needs would be much smaller.

III. GROUND WATER RESERVOIRS

The term "ground water reservoir" is widely and loosely used to denote places where ground water is accumulated under conditions that make it suitable for development and use. As there is no standard or generally accepted definition of the term, it is necessary in each use to describe the particular characteristics and limiting conditions of the reservoir. "Ground water basin," "aquifer," "water-bearing zone" (or formation), "ground water district," and "ground water area" are other terms of similarly broad and diverse usage, applied to geologic formations ranging from a few meters to thousands of meters in thickness, distributed over areas ranging from a few hectares to thousands of square kilometers, and capable of yields to wells ranging from a few liters per hour (l/h) to several cubic meters per second (m^3/s). Although it is vague, "ground water reservoir" is an excellent term for general use and is parallel in concept to "surface water reservoir" as it conveys the concept of water in storage.

A. Essential Hydrologic Characteristics

A ground water reservoir must contain formations that are not only porous, but are sufficiently permeable to yield water by gravity to springs and wells. Traditionally, ground water reservoirs, where first encountered, have been filled with water by nature, sometimes to overflowing. In contrast, surface reservoirs characteristically start out empty, and acquire stored water only after a considerable investment in study, planning, design, and construction. The addition of water to permeable formations within the unsaturated zone illustrates the use of a ground water reservoir in a similar fashion. Conversely, the reservoir would still exist even if the water were pumped out of it. Thus, the potential for storage, rather than the existence of storage, is an essential characteristic of ground water reservoirs.

Another essential characteristic of ground water reservoirs is movement of the water, both in the direction of decreasing head and generally from areas of recharge toward areas of discharge. In contrast to surface reservoirs, no reduction of velocity generally occurs in the subsurface areas designated as reservoirs; in fact, the best ground water reservoirs may have a greater velocity of movement because of the excellent permeability of the rocks; however, in comparison with stream velocities, the movement of ground water (even in the best ground water reservoirs) is so slow that practically all of it can be regarded as held in storage. The chief exception to this generalization is the flow in subterranean channels of limestone, especially in areas of karst topography or volcanic regions. Such underground streams may accept and carry the turbulent and even turbid flow of surface streams and often have equivalent velocities.

Many, if not most, ground water reservoirs contain a complicated array of permeable rock materials, which may occur in layers, conduits, compartments or lenses, each bounded by less permeable or impermeable materials. Because of the existence of many relatively permeable barriers to ground water movement, the water in some permeable rock formations is likely to be confined under artesian pressure. Unconfined water may flow laterally into rock formations where it is confined. On the other hand, confined water may leak through semi-permeable barriers and into rock formations to which it is not confined. The permeable and impermeable rocks, conduits, barriers, faults, and other geologic structural features as well as erosional and weathering effects are all parts of the geologic framework of a ground water reservoir.

Source and disposal of water vary greatly among ground water reservoirs. Where the principal inflow is from precipitation, the outflow constitutes most of the water found in streams during low rainfall periods, and the ground water reservoirs might logically be delimited by the confines of a river basin, as in the Middle Loup River basin in Nebraska. As a contrasting example, the Southern High Plains in Texas and New Mexico divide their limited natural outflow among tributaries of the Mississippi, Rio Grande, and some intrastate rivers of Texas. Some ground water reservoirs are recharged principally by a single source, but their discharge is distributed over a wide area, as in the Colorado River Delta downstream from Yuma, Arizona. Others are recharged from many streams, but have a single central area of natural disposal, such as the numerous closed basins found in the southwestern United States. Many ground water reservoirs extend across the divide separating major stream basins, and some, such as Playas Valley in New Mexico, extend across the Continental Divide.

B. Principal Types of Aquifers

The aquifers comprising the principal ground water reservoirs in the United States include both unconsolidated or partially consolidated gravel and sand and also include consolidated rocks, of which limestone, basalt, and sandstone are most noteworthy. Other rock types, such as silt and clay, shale, volcanic rocks, crystalline and metamorphic rocks, are relatively impermeable, although each of these can be sufficiently permeable to constitute an aquifer of small yield in some localities. As a general rule, the delineation of aquifers is a delineation of rock types, for the classification of rocks as sedimentary, metamorphic, or igneous, as loose or consolidated, or as soluble or insoluble quickly reveals some of their characteristics with respect to the occurrence of ground water.

1. Gravel and Sand Gravel and sand are usually of alluvial, glacial, lacustrine, marine, or aeolian deposition. They may underlie large plains and valleys or be local in character and of small areal extent.

Alluvial deposits by perennial streams may be fairly well-sorted (more uniform size of particles) and therefore permeable. The stream beds may be underlain by well-sorted sand and gravel that have been carried hundreds of kilometers from their sources. Shifting channels result in lateral distribution of these coarser stream-bed deposits, and as a result of a long period of aggradation, the entire valley floor may be underlain at various depths with sand and gravel deposits of different thicknesses. When floods overflow the stream banks, coarse material is deposited adjacent to the main stream channel, and silt and clay particles settle out in the more distant overflow areas. Mountain streams deposit the coarsest material closest to the point where the stream debouches from the canyon mouth. The stream deposits become finer with increasing distance from the point of initial deposition.

Ephemeral streams from mountains in arid areas deposit gravel and sand with much less sorting. Stream gradients and valley slopes are steep, often tens of meters per kilometer, with resulting velocities sufficient to transport large boulders. The boulders are, of course, deposited at the canyon mouth, and owing to the ephemeral nature of the flow, the remaining detrital material is deposited at the same time. This may result in sediment containing large boulders with some sand and gravel but with a clay content high enough to result in relatively low permeability.

The central and lower portions of these alluvium-filled valleys may consist of fine-grained materials deposited as stratified lake beds or playa clays. Often these clay beds are interbedded with layers of fine sand or even with stream deposits. The lower valley fill may also contain considerable soluble mineral matter in the form of caliche or other calcium carbonate salt and gypsum.

Where a stream flows into a lake or sea, the current is abruptly reduced. The coarser sediment settles rapidly, while finer materials are transported further into the body of relatively still water. Additional incursions of sand and gravel are carried to the outer edge of the previous accumulation and form rather steeply inclined beds, which are the front or face of a "delta" accumulation. Meanwhile, the silt and clay are carried still farther into the still water body and settle to form layers over the bottom of the lake or sea. As deposition continues, the delta front deposits are built progressively farther over the bottom beds. The gradient of the stream is modified by the delta formation, and sediment is deposited upstream from the original mouth of the stream. Thus, the delta grows outward by deposition of underwater beds and upwards by deposition of top beds. Deposition also occurs upstream in the original channel of the stream that provides the source material.

The deltas of larger streams are formed in the manner described in the preceding paragraph. The original underwater deposits are augmented until their surface rises above the water level. The Colorado River Delta, extending from Yuma, Arizona, across the Gulf of California, is an example of such deposition. The sand and gravel of this delta are the aquifers from which wells produce irrigation water for thousands of hectares in Mexico.

Many large ground water reservoirs are found in sand and gravel that have been deposited in extensive plains formed by parallel, coalescing, or distributary streams. Examples of these reservoirs include the Coastal Plain, formed by streams emptying into the Gulf of Mexico and the Atlantic Ocean, and the Great Plains, formed by streams draining the Rocky Mountains.

Some exceedingly permeable sands and gravel are glaciofluvial (deposited by streams flowing from glaciers). Ancient valleys that have been filled with these glacial outwash materials contain excellent aquifers, and such valleys, with the stratified deposits in glacial lakes, provide important water storage in the glaciated regions of the northern United States.

Although gravel and sand deposits are preeminent as aquifers, they constitute a very small proportion of all unconsolidated sediments. Many of these sediments are composed of a heterogeneous mixture of sands and gravels or even boulders, with predominating silt and clay that reduces the permeability so as to place these sediments in the class of aquicludes rather than aquifers. This becomes particularly noticeable in formations in which lime deposition or chemical disintegration of rock particles into clay-like materials has occurred with partial or complete closures of the pores. Both deposition and

chemical disintegration processes greatly reduce the permeability of an aquifer, and their effects are particularly noticeable with increased age and depth of formations.

2. Sandstone and Conglomerate Sandstone and conglomerate are the consolidated equivalents of sand and gravel. In these rocks, the individual particles have been cemented together, thus reducing the porosity to such an extent that they may yield little water. The best known sandstone aquifers are those which are only partially cemented. These yield most of their water from the pores between the grains, although fractures and joints may serve as collecting channels. Good sandstone wells may yield hundreds of liters of water per minute, but as a rule they yield much less than those in formations of loose sand and gravel. Extensive beds of sandstone underlie portions of the United States, and in some places comprise the only aquifer. Dakota sandstone, which derives its name from the two states where it constitutes an important source of ground water, is an aquifer of this type.

3. Limestone The limestones that are most important as ground water reservoirs are those in which a sizable proportion of the original rock has been dissolved and removed. Many very large springs issue from limestone. Of the 78 or more springs in the United States having a maximum discharge of more than 3,000 l/s, about half occur in limestone, including 27 in Florida alone. Fractures and joints in limestone also yield water in small to moderate amounts.

Most limestones are of marine origin, and many are composed largely of the fossil remains of marine organisms. If fresh water now occurs in any of these beds—that is, if the beds constitute a usable ground water reservoir—sea water must have been flushed out, and water from rain and snow must have circulated through the rock.

Solution cavities and channels in limestone may be large enough so that wells reaching them are very productive. On the other hand, rapid flow through limestone may be a health hazard if the water is polluted before it enters the limestone or receives pollution thereafter. The solution channels may be separated by considerable amounts of solid rock, so that some wells may penetrate deep into the limestone without yielding water, even though productive wells may be close by.

4. Lavas and Other Volcanic Rocks Basalt, a dark-colored volcanic rock, is an important aquifer in some sections of the United States, especially in the west, where enormous flows of lava have spread out over large areas in successive sheets of varying thickness. The hydraulic conductivity of a basalt aquifer depends on the presence of fractures, cracks, and tubes or caverns. The hydraulic conductivity of some basalt is so great that under natural conditions there is practically no slope to the water table or to the pressure surface under artesian conditions.

Other volcanic rocks, such as rhyolites and other more silicious rocks, do not usually yield water in quantities comparable to those secured from basalt. Most shallow intrusive rocks in the form of sills, dikes, or plugs are low in permeability, and many of them are impervious enough to act as barriers to ground water movement.

Crystalline and metamorphic rocks, including granite, basic igneous rocks, gneiss, schist, quartzite, and slate, originated generally at great depths below the earth's surface and are relatively impermeable. In areas where rocks of these types supply water for domestic or stock wells, the water comes chiefly from the zone of weathering or from jointed or faulted rocks. Whether the development of permeability results from weathering or fracturing, it is likely to be greatest near the surface and progressively less with increasing depth.

C. Principal Types of Basins

Closed basins in the southwestern United States are generally rimmed by mountain ranges and have centripetal drainage, which forms a lake in the lowest part of the basin whenever runoff from precipitation is sufficient; however, many of these basins are in arid regions where there is no stream flow except after occasional storms, and where water principally moves from the bordering mountains toward the lowlands as ground water. In such basins, the mountains may constitute ground water divides as well

as topographic divides. The low part of the basin constitutes a sink from which ground water is discharged by evapotranspiration. Thus, a ground water basin exists which is coextensive with a drainage basin; however, this may be the exception rather the rule. Many ground water basins do not conform with surface drainage basins, either in area or in pattern of flow.

If ground water basins are to be subdivided, their sub-basin boundaries would logically be the relatively impermeable barriers that make such subdivision necessary. The Santa Ana River valley in southern California contains a ground water basin that is an excellent example of natural compartmentalization by faults and other barriers. Pumping draws more heavily upon some compartments than others, revealing the skeleton of less permeable barriers and the varying rates at which water can flow through, around, or over them. For management purposes, a ground water basin may have to be subdivided using barriers to flow as well as political boundaries. Such approaches are not often desirable but may be required for practical reasons.

The term "basin" may also express contrasting degrees of confinement as, for example, in an artesian basin. Particularly in the valley fill of arid basins, where there are likely to be large variations in the permeability of sediments tapped by a single well, some water may be confined, and some unconfined in one season and confined in another, while in many other areas water can be described as "semiconfined." In broad perspective, a ground water basin containing water under confined conditions is not a complete and self-supporting hydrologic unit, for it generally does not receive water by downward movement through the overlying confining bed. Even if the confining bed were moderately permeable, the artesian pressures would oppose such movement; however, if the artesian head is below ground level or below the free water table, as it is in many areas, the artesian aquifer may receive recharge through the confining bed. This phenomenon happens in many regions. An artesian basin, therefore, is replenished in areas where the water table is above the artesian head, primarily in permeable recharge areas where the confining bed is absent or somewhat permeable. Although it is important to differentiate between confined and unconfined waters (because of the differences in ground water hydraulics) it should be recognized that an artesian basin is actually only a part of the complete underground system necessary for yielding perennial supplies to wells or springs.

In a free or unconfined ground water basin, the water table represents the height to which the basin has been filled. Changes in the water level in wells that tap the basin are direct indications of the change in the volume of water in storage. A summation of the products of the changes in water level, the average specific yield of the aquifers in the zone of water level change, and the area dewatered or recharged (as the case may be) gives the change in the quantity of stored water.

The water table slopes in the direction of ground water flow, from areas of recharge at higher elevations to areas of discharge at lower elevations. If the water moves into confined aquifers, the boundary between the unconfined and the confined ground water basins is logically the edge of the confining bed. In many ground water reservoirs, the terms "forebay" and "outcrop" have been used to designate a free or unconfined ground water area that serves as recharge area to an artesian or confined area (see Fig. 5.1).

In a confined ground water basin, the aquifer is completely saturated and confined by overlying impervious or relatively impervious materials. Changes in the potentiometric surface, determined by water level or pressure measurements in wells that tap the confined aquifer, represent changes of pressure in the aquifer. Such changes in pressure may reflect changes in head of the recharge, outcrop, or discharge area. They may also represent changes in atmospheric pressure, but the largest recorded changes have resulted from pumping in the vicinity of the observation wells.

IV. THE SUBSURFACE MEDIUM

The medium within which ground water occurs is porous, contains secondary openings such as fractures or bedding planes, or both. The availability, movement, and quality of ground water frequently depend directly upon the type of medium within which it is found.

A. Porous Medium

Some openings in geologic materials are "primary" in the sense that they were formed contemporaneously with the materials, e.g., the spaces between individual particles of gravel, sand, clay, volcanic tuff, and breccia, or the gas cavities formed during cooling of volcanic rocks. In sedimentary rocks, porosity depends upon the shape, arrangement, and degree of sorting of the particles. For example, spheres of uniform size have a porosity ranging from 26% to 47.6%, depending upon their arrangement. In sediments, except for some limestones, the porosity is generally greatest at the time of deposition and is then reduced by compaction (closer packing of particles) and cementation (deposition of mineral matter around the particles). Even though the process of consolidation may also include the development of secondary openings by fracturing, the net product of consolidation is generally a reduction in porosity. A very large proportion of the water yielded by wells comes from primary openings in unconsolidated or semiconsolidated sediments, chiefly gravel and sand. In several localities, water withdrawal has resulted in further compaction of fine-grained materials, as shown by subsidence of the land surface, and has thus reduced the overall porosity in subsided zones.

B. Medium with Secondary Openings

Fissures, fractures, bedding planes, and solution cavities are examples of "secondary" openings in geologic materials through which ground water can move. Such openings are created after the particular geologic material has been emplaced and are caused by cooling, folding, faulting, or chemical action. Fractures usually are larger and more numerous near the surface, because increasing pressure with depth tends to minimize fracture aperture size. Solution cavities are common in limestone and other soluble rocks. Cavities may result from ground water solution of the rock minerals, but fractures may have provided the original path enabling the water to make contact with the minerals. Weathering is an important agent in formation of secondary openings in crystalline and cemented sedimentary rocks in near-surface exposures. Plant roots and animal and insect borings may also create secondary openings in rocks and soils.

1. Porosity Porosity is the ratio of the total volume of pores in rock or soil material to the total volume of the material. The porosity of most rock materials ranges from 1% to 50% by volume. Typical value ranges of porosities for some rock materials are given in Table 5.1. In aquifer material, porosities less than 5% are considered small, those between 5% and 20% are considered medium, and those greater than 20% are large. Pores may range in size from large caverns to the microscopic spaces between crystals. Pores are classified as "subcapillary" (so small that water is held primarily by molecular force), "capillary" (small enough to hold water by surface tension), and "supercapillary" (large enough to allow gravity drainage or to yield water to wells).

Porosity is an inherent property of the rock material independent of the presence or absence of water, which can be determined by laboratory measurements (Davis, 1969); however, for ground water studies, total porosity is less important than effective porosity, or the interconnected pore space that can be drained by gravity (Dagan, 1967). If a unit volume of material is saturated (all pore space filled with water) the volume of water that is yielded by gravity drainage is the specific yield, and the volume retained by molecular forces is termed the specific retention. Both values are expressed as percentages of the total volume of rock and together constitute the total porosity of the given rock material.

The porosity of unconsolidated rock materials *in situ* is frequently changed by human activities. For example, in several localities, extraction of ground water has resulted in consolidation of sediments, primarily of clay deposits, as demonstrated by subsidence of the land surface (Poland, 1969, 1981; Poland and Davis, 1969). Also, collection of samples by drilling or some types of core barrels may disturb soils from their natural *in situ* structure. Accordingly, laboratory tests on such disturbed samples may show porosity and specific yield values that incorporate some error (Johnson, 1967).

TABLE 5.1. Typical Value Ranges for Porosities and Specific Yield for Various Aquifer Materials.

Aquifer materials	Porosity (in %)	Specific yield
Clay	45-55	1-10
Fractured Crystalline Rock	0-10	
Glacial Till	10-25	
Gravel	25-40	15-30
Limestone	1-20	0.5-5
Sand	25-40	10-30
Sand and Gravel	10-35	15-25
Sandstone	5-30	5-15
Shale	0-10	0.5-5
Silt	35-50	
Vesicular Basalt	10-50	

2. Specific Retention The specific retention of a rock or soil is the ratio of the volume of water which the rock or soil, after being saturated, retains against the pull of gravity to the volume of the rock or soil. Ideally, the definition implies that gravity drainage is complete; however, the quantity of water held in pores above the water table during gravity drainage depends upon particle size, distance above the water table, time of drainage, and other variables. Lowering of the water table and infiltration occur over such short periods of time that gravity drainage is rarely or never complete. Nevertheless, specific retention is a useful though approximate measure of the residual moisture-holding capacity of the unsaturated zone in the region above the capillary fringe.

3. Specific Yield The specific yield of a rock or soil is the ratio of the volume of water that the saturated rock or soil yields by gravity to the volume of the same rock or soil. The definition implies that gravity drainage is complete. In the natural environment, specific yield is generally observed as the change that occurs in the amount of water in storage per unit area of unconfined aquifer as the result of a unit change in head. Such a change in storage is produced by the draining or filling of pore space and also depends upon particle size, rate of change of the water table, time, and other variables. Hence, specific yield is only an approximate measure of the relation between storage and head in unconfined aquifers. It is equal to porosity minus specific retention. Typical value ranges for specific yield of some rock materials are given in Table 5.1.

Specific yield is an important concept in ground water hydrology (Lohman, 1972; Youngs, 1969). The average specific yield of saturated sediments, multiplied by the total volume of those sediments, gives the volume of water contained in those sediments that can be recovered by gravity drainage (Johnson, 1967). No time factor is included in the definition of specific yield, but many experiments have shown that draining is initially most rapid, decreases with time, and may continue for months or even years (Prill et al., 1965). Fine-grained materials have a lower specific yield than coarser materials, even though their porosity may be greater, and the drainage process is likely to continue for a much longer period of time than for coarse materials. The difficulty in individual determination of specific yield and the problem of extrapolating the results to the volume and variety of rocks in a ground water reservoir lead to the conclusion that excellent opportunities exist for refining existing techniques and developing new ones for determining usable storage of a ground water reservoir. Some techniques for determining specific yield (Johnson, 1967) and their limitations are:

(a) Draining long columns of saturated materials in the laboratory (Johnson et al., 1963). For accuracy, it is essential to reproduce the natural conditions as specific yield is related not only to the size and uniformity of the material, but also to the arrangement of the grains. The limitation to this method is that it is very difficult to obtain undisturbed samples that truly represent natural conditions.

(b) Saturating, in the field, a considerable body of material situated above the water table and allowing it to drain downward naturally. After drainage, the moisture content of the material is determined by collecting core samples, using a nuclear moisture meter or other techniques.
(c) Measuring moisture content retained immediately above the capillary fringe after the water table has receded an appreciable distance, as is common in summer and autumn. This technique is essentially the same as the field saturation and draining method.
(d) Ascertaining the volume of sediment drained by heavy pumping and keeping a record of the quantity of water that is pumped. This method requires accurate records of water pumped and the magnitude and extent of the decline in water table. It is accurate only if there is no recharge to, and no other discharge from, the aquifer during the period of pumping.
(e) Ascertaining the volume of sediments saturated by a measured amount of precipitation or seepage from one or more streams. This method requires accurate records of water recharge and the magnitude and extent of the rise in water table. It is accurate only if there is no discharge from, and no other recharge to, the aquifer.
(f) Making indirect determinations with small samples in the laboratory by the application of centrifugal force (Johnson et al., 1963). This technique is also used to determine moisture equivalent.
(g) Making particle-size analyses and estimating from there the specific retention and the specific yield. The accuracy of the estimate depends upon sufficient statistical data showing the relation of specific yield to particle-size distribution based on actual test data. In certain areas the sediments may be composed of materials that have enough similarity to permit the development of reliable relationships. In a number of instances, this method has been done fairly successfully.

4. Permeability, Hydraulic Conductivity, and Transmissivity The permeability of a porous medium describes the ease with which a fluid will pass through it, and indicates its capacity to transmit water under a head differential. No rock material is absolutely impermeable in the strictest sense of the word; permeability is a relative term. Rock materials of low porosity characteristically have low permeability as well, but high porosity does not necessarily produce high permeability. Rock materials with isolated interstices, such as some volcanic rocks, may be quite porous, but necessarily have very little permeability. It is common to find rocks that have very little permeability in one direction, such as the vertical axis, and a much higher permeability in another direction, such as along the horizontal axis; i.e., they are anisotropic with respect to permeabilities.

The permeability, k, is dependent only on the properties of porous media and is independent of the fluid properties that govern the flow. Earlier researchers have defined k as the specific permeability proportional to the square of representative grain diameter of the medium, or the constant of proportionality depending upon porosity, packing, size distribution, and shape of grains (Reeve, 1957). Several formulas have been developed that include variables designed to express the properties of the porous media, but they have encountered difficulty due to the complexity of variations in porous media (Dagan, 1967). The specific permeability is commonly expressed in "Darcy" units, first introduced in the petroleum industry. These units depend upon the medium but not upon the fluid that flows through that medium. The U.S. Geological Survey has adopted the term "intrinsic permeability" for k for use in ground water studies.

Hydraulic conductivity, K, in consistent units has been adapted to replace permeability in the inconsistent units gpd per ft^2 (gallons per day per square foot). Hydraulic conductivity is the flow of a unit volume of water per day under a unit hydraulic gradient through a unit cross-sectional area at prevailing temperatures. In common units, hydraulic conductivity is expressed as centimeters per second or meters per day, under a unit hydraulic gradient. Table 5.2 gives the relationship of the various units used for hydraulic conductivity and permeability. The relationship between hydraulic conductivity and permeability is:

$$K = \frac{k\rho g}{\mu} \tag{5.1}$$

TABLE 5.2. Hydraulic Conductivity–Permeability Conversion Factors.[a,b]

Unit to be converted (1)	Centimeters per second (2)	Feet per day (3)	Feet per year (4)	Darcy (5)	Meinzer (gpd per sq ft) (6)
Centimeters per second	1	2.835×10^{3}	$1{,}0348 \times 10^{6}$	1.033×10^{3}	2.12×10^{4}
Feet per day	3.53×10^{-4}	1	365	3.64×10^{-1}	7.48
Feet per year	9.67×10^{-7}	2.74×10^{-3}	1	9.99×10^{-4}	2.05×10^{-2}
Darcy	9.68×10^{-4}	2.75	1.001×10^{3}	1	20.50
Meinzer (gpd per sq ft)	4.72×10^{-5}	1.34×10^{-1}	48.8	4.88×10^{-2}	1

[a] Multiply unit at left by number in column to get unit at top of column. All units based on temperature of 60°F or 15.6°C.
[b] From Johnson, A.I., "Application of Laboratory Permeability Data," U.S. Geological Survey open-file report, 1963.

where K is the hydraulic conductivity, k is the permeability, ρ is the density, g is gravity, and μ is viscosity. K has the units of length divided by time (L/T) and k has the units of length squared (L^2).

Typical value ranges of hydraulic conductivity, K, for various rock materials are given in Table 5.3.

Rock materials within the unsaturated zone have less hydraulic conductivity than do the same rock materials if they are saturated. This relationship is logical because the movement of water is impeded by the air that occupies some of the available void space. Generally, the lower the amount of water in the void space, the lower the hydraulic conductivity down to a threshold point (approximately representative of the moisture content at specific retention) where there is essentially no liquid flow.

Transmissivity, as defined and used by the U.S. Geological Survey, is the product of the hydraulic conductivity multiplied by the thickness of the saturated portion of the aquifer. It represents the quantity of water that flows through a unit width of the saturated portion of the aquifer under a unit hydraulic gradient and at prevailing water temperature.

V. MOVEMENT OF WATER

Movement in the unsaturated zone is usually vertical but may have a significant horizontal component because of the effect of permeability on the rate and direction of movement. Thus, the unsaturated zone may include layers of materials having contrasting permeability or some layers in which horizontal permeability greatly exceeds vertical permeability.

TABLE 5.3. Typical Value Ranges for Hydraulic Conductivity, K, for Various Aquifer Materials.

Aquifer materials	Hydraulic conductivity (m/day)
Clay, Unweathered Marine	10^{-4}—10^{-7}
Fractured Igneous and Metamorphic Rocks	10—10^{-3}
Glacial Till	10—10^{-7}
Gravel, Fine to Coarse	10^{5}—10^{2}
Karst Limestone	10^{3}—10^{-1}
Limestone, Unjointed Crystalline	10^{-1}—10^{-5}
Sand, Fine to Coarse	10^{3}—10^{-2}
Sandstone, Cemented and Unjointed	10^{-2}—10^{-5}
Sandstone, Friable	1—10^{-3}
Shale	10^{-4}—10^{-8}
Silt or Loess	1—10^{-4}
Tuff	3—3×10^{-5}
Unfractured Igneous and Metamorphic Rocks	10^{-5}—10^{-9}
Vesicular Basalt	10^{3}—10^{-3}

The movement of water in the saturated zone ordinarily has a large horizontal component in the direction of decreasing head; however, movement in the saturated zone, even in unconsolidated granular materials is seldom uniform in direction or in velocity. Water may move around deposits of impervious material or follow the stratified courses of the more permeable material. Water follows the path of least resistance and moves with less restriction through sections which have the higher permeability.

A. Velocity

When water moves through a uniformly permeable material at very low velocities, all molecules move in more or less parallel lines. As the velocity increases, a value termed "higher critical velocity" is reached, at which time the streamline or laminar flow ceases and movement becomes turbulent. When the velocity of turbulent flow is gradually reduced, a velocity known as "lower critical velocity" is reached at which the flow again becomes laminar. At velocities intermediate between the higher and lower critical velocities, the flow is unstable. Although turbulent flow may occur in large joints, fractures, or in solution openings in limestone, ground water in granular materials ordinarily moves at laminar velocities. Turbulent flow in sand and gravel aquifers is virtually nonexistent, except possibly at the point of water entry into some pumped wells.

The velocity of laminar flow follows Darcy's law, which states that velocity varies directly with the hydraulic gradient, or

$$v = K\frac{\Delta h}{l} \tag{5.2}$$

in which v is velocity, Δh is the difference in head between two points in the path of movement, l is distance along the path of movement, and K is hydraulic conductivity. For water moving in the zone of saturation, the factor $\Delta h/l$ is the hydraulic gradient.

The velocity of flow is influenced by the viscosity of the water, which varies inversely with the temperature. For example, an increase in water temperature from 4°C to 32°C approximately doubles the velocity under the same hydraulic gradient. Although the temperature of ground water is fairly constant throughout the year, its temperature increases with depth. In areas where shallow ground water is derived by rapid infiltration from a recharge source, the ground water temperature may vary in response to the changes in surface water temperature. In some localities, this factor is of practical importance.

Hydraulic gradients and, by inference, the directions of ground water movement are commonly portrayed on two-dimensional maps. The position of the water table is indicated by the static water level in wells tapping unconfined water. The form of the water table is commonly shown on maps by contour lines connecting points of equal elevation of the water table. As with the contours on topographic maps, the direction of maximum gradient (and therefore the horizontal component of ground water movement) is typically perpendicular to the contours. As the direction of water movement is down-gradient, the areas of recharge constitute highs on the maps. Influent streams, for example, commonly form ground water ridges. Conversely, areas of discharge are lows on the map, including valleys under effluent streams (where the contours bend upstream). Closed depression contours are common in the vicinity of discharging wells. In contrast to inferences that may be drawn from topographic maps concerning stream velocities, the spacing of ground water contours may be related inversely to the rates of water movement. Commonly, a steepened water table slope indicates decreased hydraulic conductivity or thinning of the aquifer and a flattened slope indicates increased permeability or thickening of the aquifer.

Water levels in wells tapping a confined aquifer are the basis for constructing contour maps or profiles of a potentiometric, or pressure, surface for the aquifer. Such a surface indicates the potential energy head of the water in the aquifer. A map of the potentiometric surface can be the basis for inferences as to hydraulic gradients and recharge and discharge areas, similar in concept to those drawn from water table maps; however, an isolated closed high, which might be interpreted as a point source of water, is more

likely to indicate a well that taps a different aquifer. In addition, particularly in confined aquifers, a closed depression may be an indication of upward movement through confining deposits rather than horizontal movement into that location. Because the map indicates only the horizontal component of movement, it cannot portray actual movement through an aquifer that has the structure of a syncline (inverted siphon), for instance.

In any confined aquifer capable of perennial supply to wells, the confinement must end somewhere to permit replenishment of the supply from precipitation, stream flow or other surface source, or the confining layer must leak sufficiently to provide a source of recharge water to the wells. A map that shows the hydraulic gradients throughout a complete hydrologic unit might portray one or more potentiometric surfaces in some areas and a water table in other areas.

Dyes, chemical ions, and radioactive tracers have been used to measure the direction and velocity of ground water movement through porous media with varying degrees of success. Base exchange of cations, if the porous media have this capability, may seriously impair the usefulness of certain chemicals. As summarized by Koufman and Orlob (1957), the ideal tracer would be one that is capable of detection in very low concentrations, does not react with injected or displaced waters, is not removed from solution or acted upon by the porous medium, is readily obtainable and economical, and is safe for use with potable water sources.

B. Quantity of Flow

If the velocity of movement has been measured, it may be used in computing the quantity of flow. This quantity of water discharged in a unit time, Q, is equal to the product of the average pore water velocity, V_{eff}, the average effective porosity, η_e, of the media through which the water flows, and the gross cross-sectional area, A, of the aquifer:

$$Q = V_{eff}\, \eta_e\, A. \tag{5.3a}$$

Effective porosity is the portion of interconnected pore spaces through which water can move. The effective porosity and cross-sectional area may be estimated roughly from the analysis of material encountered in well logs and geological studies.

Alternatively, the quantity of flow may be determined by the product of the hydraulic conductivity, K, the hydraulic gradient, I, along the flow direction, and the gross cross-sectional area, A, through which the water flows:

$$Q = KIA. \tag{5.3b}$$

This is essentially a restatement of Darcy's law.

C. Flow of Ground Water in Three Dimensions

While Darcy's law is generally applied to the determination of average horizontal velocity of flow resulting from head differential over an appreciable distance, it is equally applicable to water flow in any direction occurring in any part of the aquifer water mass.

In general form, Darcy's law may be written as

$$\nu = -K\frac{\partial h}{\partial s} \tag{5.4}$$

in which ν is Darcy velocity (flux), K is hydraulic conductivity in the s direction (∂s = distance changes), and h is the total energy or head at the point under consideration (∂h = head changes). Flow along the

three principal coordinate axes (anisotropic) may consequently be described as (Kashef, 1976; Reeve, 1957)

$$v_x = -K_x \frac{\partial h}{\partial x}, \quad v_y = -K_y \frac{\partial h}{\partial y}, \quad v_z = -K_z \frac{\partial h}{\partial z}. \tag{5.5}$$

In isotropic aquifers, $K_x = K_y = K_z$. Continuity under conditions of steady flow without change in time and density is expressed by

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_2}{\partial z} = 0 \tag{5.6}$$

which, in view of the hydraulic conductivity K, becomes

$$\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} + \frac{\partial^2 h}{\partial z^2} = 0 \tag{5.7}$$

which is a form of the well-known Laplace equation for ground water flow through an isotropic homogeneous aquifer under steady-state conditions.

D. Flow Nets

Flow nets (Cederegren, 1967; Todd, 1980) are graphical representations of the pattern of ground water flow in both vertical and horizontal projections. The net is composed of two families of curves, one of which represents the flow lines, while the other represents the variation in head in the aquifer or flow cross section (potential lines). The mathematical basis for the flow net can be derived from the Laplace equation, the third term eliminated with the assumption of two-dimensional flow.

Flow nets are two sets of curves or lines for isotropic media that form an orthogonal pattern of squares or rectangles. As most ground water situations involve curvilinear paths of flow, the form of net tends to assume a curvilinear pattern; however, the points of flow and head intersection or equipotential lines for isotropic media should be right angles, and the mean distance between opposite sides of the resulting rectangles should be equal for homogeneous media. Repeated divisions of such orthogonal patterns should produce subdivisions that approach the shape of true squares.

The use of flow net principles facilitates the solution of many problems involving water movement through either confined or unconfined aquifers. Flow lines drawn perpendicular to water table or potentiometric contours indicate the general direction of ground water flow in a specific area (see Fig. 5.2). Knowing the direction of ground water movement has become increasingly important in considering the danger of contaminating ground water supplies. A flow net can indicate the direction of contaminant travel. Areas of recharge and discharge can also be defined with a flow net map. Flow lines diverge in areas of recharge and converge in areas of discharge. In areas of uniform ground water flow, wide contour spacing (flat gradients) indicate areas of higher hydraulic conductivity than those areas with narrow contour spacing (steep gradients). Flow nets can be helpful in locating water wells by indicating areas with a good source of supply (recharge) and favorable hydraulic conductivity.

VI. GROUND WATER BASIN YIELD CONCEPTS

Water is universally classified as a renewable resource. Renewal of water supplies is inherent in the hydrologic cycle, and it is obvious that precipitation replenishes soil moisture and streamflow. In concept, ground water is also recognized as a generally renewable resource because virtually all usable supplies

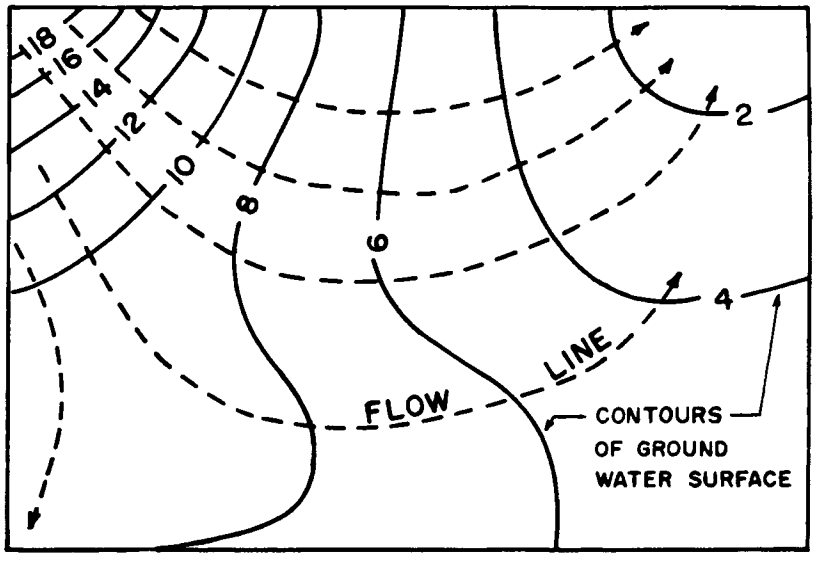


Figure 5.2—Typical Flow Net

are atmospheric in origin. Detailed studies of many ground water areas have shown conclusively that ground water supplies are recharged by precipitation. There are no known examples of fresh ground water resource areas that are isolated from such possible replenishment, even though the replenishment may have been ancient.

Nevertheless, quantity of ground water replenishment is limited. The average annual precipitation on the land area of the United States is about 6×10^{12} m^3. Of this amount, a portion evaporates, a portion is used in transpiration, and a portion runs off to the oceans. Only a small fraction of the 6×10^{12} m^3 is available for ground water replenishment. Nearly 2.5×10^{14} m^3 of fresh ground water are estimated to reside within 1,000 m of the land's surface. Apparently, the quantity of ground water that can be considered available as a renewable resource is limited by the quantity of precipitation replenishment. The replenishment is a small fraction of the total precipitation per year plus the reduction of nonbeneficial losses from the ground water reservoir.

A perennial supply of water can be assured by not withdrawing more water than is available over a long term. This concept holds true for either ground water or surface water projects. Determining the perennial yield from ground water sources is more complicated than determining yield from surface water projects because it is difficult to differentiate replenishable ground water from accumulated stored ground water.

A. Perennial Yield

The perennial yield of a ground water reservoir is defined as "the maximum quantity of water which can be withdrawn annually from a ground water supply under a given set of conditions without causing an undesirable result." The words "undesirable result" are key to the definition. Adverse or deleterious side effects of ground water development, such as sea water intrusion, land surface subsidence, salt water coning, and others are included. As the definition states, perennial yield is determined for a specified set of operating conditions. Any change in conditions, such as changes in economics, land use, or importation of new water supplies requires calculation of a new yield. The term "safe yield" is often used instead of the preferred "perennial yield."

The concept of perennial yield of a ground water reservoir stems from the principle of renewable resources, as shown in the hydrologic cycle, and is based on an analogy to surface reservoir operation. The perennial yield from a designated ground water reservoir can be achieved if the artificial discharge

by wells are patterned to reduce discharge from the reservoir and to induce recharge in an equal amount, and if storage is utilized only to provide regulation of the fluctuating inflow to meet the demand on the wells. In practice, three complicating factors must be taken into account.

First, if the water table is close to the ground surface, either initially or during the operating period, some recharge to the water-bearing formation may be rejected. As water cannot be added to storage when a reservoir is full, the excess water must be evaporated or wasted. Development of a previously unused or underused aquifer usually results in increased recharge from surface streams and precipitation or a reduction in ground water discharge to a surface stream. Therefore, use of an aquifer replenished by meteoric water may diminish surface runoff so that the yield from an aquifer reduces the water available from surface sources. Jenkins (1968) discusses the theory of stream depletion by the operation of wells and provides sample calculations. Conover (1954, 1961a, 1961b) portrays the effects of ground water use on surface water and the benefits of coordinated surface water–ground water use.

Second, an aquifer contains a supply of water that can be mined, regardless of whether the aquifer is replenished. The perennially available water supply therefore may be substantially augmented for a very long time indeed—50 years of mining is not an unusual circumstance.

Third, the patterns of distribution, rates, and volumes of pumpage from wells in a developed ground water reservoir commonly differ from the ideal pattern for obtaining the maximum perennial replenishment. The maximum yield then may depend on the practicability of locating or relocating wells rather than on any characteristic of the natural resource. Pumping from a ground water reservoir in quantities greater than the perennial yield may modify one or more of the items in the hydrologic equation, thereby reducing or increasing the yield over time.

The perennial yield of an aquifer, in some instances, can be substantially augmented by engineering controls. For example, more water can be made available through artificial recharge by spreading basins or injection wells or lowering ground water levels to reduce evapotranspiration to capture rejected recharge or surface water from streams. The concept of changing the perennial yield of an area requires applied water management, including the services of specialists in hydrology, geology, and engineering for evaluation. The amount of ground water production quantified by the phrase "perennial yield" is fixed at any point in time only in the sense that no more money may be available for engineering construction, or that legally no more water may be obtained from any source. Should these constraints be changed, for example, by the importation of water and the utilization of underground storage, the perennial yield could be increased. Under such circumstances, the concept of a specific quantity for perennial yield is not a fixed amount but a changing quantity depending on the set of conditions applicable to the aquifer area. It involves consequences of pumping, such as economics, environmental effects, change in quality, and deformation of aquifers, which are deemed acceptable or safe. In contrast, overdraft involves any factor that may be termed unsafe; it involves more than just a continuing decline in storage or ground water mining.

B. Mining Yield

Water can be mined from an aquifer as long as the water table or potentiometric surface is consistently lowered in elevation over a period of years as a result of a withdrawal of ground water in excess of the induced recharge plus the reduced ground water discharge. This assumes that proper allowances are made for climatic variations.

In many areas—the High Plains areas in Colorado, Nebraska, Kansas, Texas, and New Mexico, for instance—the replenishment of ground water is not only negligible but additional recharge cannot be induced by lowering the water table. Further, the natural discharge is so diffuse, small, and distant from the centers of pumpage that, in practical terms, the discharge is minimally affected by the pumpage. In this sense, the water essentially is being mined. Stored water in this area was estimated by the Texas Water Development Board (Texas Board of Water Engineers, 1951) as about 3.54×10^{11} m^3, of which possibly 2.22×10^{11} m^3 to 2.47×10^{11} m^3 was considered recoverable. The quantity available is presently

a 40-year or 50-year supply at the 1969 rate of mining, with no side effects such as land surface subsidence or salt water encroachment to be considered. Water management in this instance involves keeping track of ground water pumpage and the quantity remaining in the aquifer, leading an educational campaign to reduce use and waste of water and conducting a political effort to import a new, replacement water supply. In addition, more economic uses of water, such as industrial rather than agricultural, and arbitrary reduction in the rate of pumpage may be of value. The establishment of well-spacing regulations also could be useful.

Aquifers are usually variable in thickness and depth. Water-mining operations are, consequently, variable in rate and duration, depending on the location of wells in the aquifer. Property owners whose wells tap the lowest elevations of the aquifer normally enjoy wells possessing the largest yields and longest lives. Those whose properties overlie thin or shallow aquifers find that their wells lose capacity more quickly.

Under identical conditions of hydrogeology, different results would be obtained if the same quantity of water were to be used for industrial processes, cooling tower makeup, or drinking water rather than for agriculture. The water deposit would be minable for a very long time in such cases. Therefore, even under the classic conditions existing in the High Plains area of the Central United States, the life of a water mining operation cannot be predicted from engineering data alone. A management plan including regulation of well-drilling, well-spacing, and pumpage rates would provide an estimate of the useful life of the deposit and would assist in the development of the resource by providing a definite life for investment purposes. Under the circumstances existing in mined areas, public management of ground water resources might be concentrated on education, determination of the rate of ground water decline, and the depletion of the ground water reserve for tax purposes.

C. Sustained Yield

The term "sustained yield" is most useful when applied to well fields in situations that Thomas (1957) has termed "water-course aquifers." Such aquifers generally underlie the floodplain of a major river system and are in hydraulic contact with the water in the river. The sustained yield is the minimum rate of pumpage sustainable under all conditions of river discharge (and temperature) by a specified well field that taps the alluvial aquifers.

Sustained yield is not a fixed quantity. It can be changed by water management efforts even if the results of constructing and operating injection wells, spreading grounds, or recharge pits to improve recharge conditions are excluded. Additional wells can be constructed and operated while evapotranspiration is reduced, and the yield of the system (assuming the permeability and area of the river bed is not restricting the natural recharge) can be increased, albeit in progressively smaller, more expensive, steps (Conover, 1954).

D. Deferred Perennial Yield

Another ground water basin yield concept is deferred perennial yield. In a ground water basin or district, the quantity of water available for all uses might be limited by legal restrictions or the price of water; however, it is possible that a significant quantity of water can be obtained from aquifer storage. The withdrawal of this water might be cheaper than importing an equal quantity from outside the basin. Lowering the potentiometric surface often produces no undesirable side effects.

Therefore, extractions at a rate greater than the ultimate steady-state supply of ground water may be acceptable. For an initial period of years, the pumpage may be far greater than the anticipated steady-state supply of water. The economic use of water withdrawn from storage frequently results in a strengthened economy capable of importing water from a greater distance and at a higher cost than was originally anticipated. The danger in using this approach is that the proposed new source of water may not be available in the future due to changing social and economic values. Also, managers must be aware of possible damage to natural habitats that may be caused by extensive lowering of ground water levels;

CHAPTER 6

RUNOFF, STREAM FLOW, RESERVOIR YIELD, AND WATER QUALITY

I. INTRODUCTION

Runoff, or surface runoff, refers to all the waters flowing on the surface of the earth, either by overland sheet flow or by channel flow in rills, gullies, streams, or rivers. Stream flow refers to the flow in natural streams. Runoff and stream flow are continuous processes by which water is constantly flowing from higher to lower elevations by the action of gravitational forces. In this way, overland flow concentrates into small streams, which in turn combine to form larger streams and rivers. Eventually, rivers flow into oceans completing the hydrologic cycle.

Runoff is usually expressed in terms of either volume or flow rate. The usual units of runoff volume are cubic meters, cubic feet, or acre-feet. Flow rate or discharge is usually expressed in cubic meters per second (m^3/s) or cubic feet per second (cfs). Discharge at a cross-section or gaging station usually varies in time; therefore, its value at any time is the instantaneous or local discharge. Instantaneous values of discharge can be integrated over a period of time to give the runoff volume for the entire period.

Runoff is also expressed in terms of depth units, by dividing the runoff volume by the catchment/watershed/basin area to obtain an equivalent spatially averaged runoff depth. In certain applications of flood hydrology runoff can also be expressed as: a) peak discharge per unit drainage area ($m^3/s/km^2$; cfs/mi^2), b) peak discharge per unit runoff depth ($m^3/s/cm$; cfs/in), and c) peak discharge per unit drainage area per unit runoff depth ($m^3/s/km^2/cm$; $cfs/mi^2/in$).

This introductory section sets the stage for the remainder of the chapter. It contains two subsections: 1) Description of Physical Processes and 2) Variability of Runoff. Description of Physical Processes seeks an understanding of the runoff processes as a framework for the detailed calculations that will follow. Variability of Runoff describes the aspects inherent to the study of runoff, namely its local, spatial, temporal, seasonal, regional, and geographical variability.

A. Description of Runoff Process

Surface runoff comprises all the waters flowing on the surface of the earth in response to precipitation. Surface runoff eventually goes on to constitute stream flow; however, stream flow comprises both surface and subsurface flow. The subsurface flow component of stream flow consists of exfiltration from interflow and ground water flow.

Surface runoff occurs when water originating in precipitation (rainfall and snow) flows freely on the surface of the earth, driven by gravitational forces. Typically, the first manifestation of surface runoff is

overland sheet flow. In overland flow, water flows under laminar or mixed laminar/turbulent flow conditions as sheet flow on the surface, usually at very small depths and velocities.

Given the topographical irregularity of the earth's surface, overland flow soon concentrates into rill flow; in turn, rill flow concentrates into gully and stream flow. Rill flow is the flow in small rivulets. These feature greater depths and velocities than the adjacent or contributing overland flow. The distinction between rills and gullies is primarily one of scale; rills can be obliterated by common agricultural practices. Gullies, on the other hand, are a more permanent feature of the landscape. The distinction between gullies and streams is largely one of age. Gullies are young in a geologic time frame while streams are older.

1. Overland Flow Generation Three distinct physical processes and models have been developed to explain the process of overland flow generation. These are: a) Hortonian overland flow, b) saturation overland flow, and c) overland flow due to surface crusting. In Hortonian overland flow (named after Robert E. Horton, the renowned American hydrologist), the excess rainfall intensity (i.e., the rainfall intensity minus the rate of infiltration) flows on the ground surface first as overland sheet flow, later as rill and/or gully flow, eventually concentrating into streams and rivers. This type of overland flow is typical of highly intensive rainstorms on arid and semiarid catchments, which often lack enough vegetation cover (and associated litter and shallow root environment) to retain moisture for extended periods. As soon as the rainfall intensity exceeds the capacity of the soil to absorb water, the flow runs on the ground surface toward lower elevation.

Saturation overland flow occurs when the soil profile is saturated. In this case, overland flow occurs—regardless of the difference between rainfall intensity and infiltration rate—due to the soil's inability to absorb any more water. It follows that saturation overland flow occurs when long-duration general storms cover large basins having relatively shallow soil mantles. Typically, saturation overland flow occurs as a result of a storm couplet (i.e., a sequence of two storm events). The first storm saturates the soil mantle; the second storm—following on the wake of the first—goes on almost entirely to constitute overland flow and surface runoff.

The third process of overland flow and surface runoff generation is caused by the phenomenon of surface crusting. This mode of surface runoff generation occurs when highly intensive storms fall on arid and semiarid catchments devoid of protective vegetation cover. The excess rainfall intensity leads to Hortonian flow and, depending on the upper soil layer's texture and structure, to splash erosion and associated entrainment of silt-sized particles in the flow. The entrained sediments are transported downwards, segregated by size, and eventually deposited on the surface to form an impermeable crust. Once the surface crust is developed, infiltration is reduced to negligible amounts, and essentially all subsequent rainfall is converted into surface runoff. The formation of a surface crust is, by nature, highly random and dependent on a suitable combination of climate, storm type, vegetation cover, soils, and topographic relief.

2. Stream Flow Generation Stream flow generation differs from overland flow generation in that it includes both surface and subsurface runoff. Classical hydrology defines stream flow in terms of three components: a) surface runoff, b) interflow, and c) ground water flow. Notwithstanding classical hydrology, recent theories of stream flow generation have emphasized the timing, rather than the path, of the components of stream flow. Two components are recognized under this framework: a) quickflow, consisting of surface runoff (overland, rill, and gully), fast interflow, and rain falling directly on the channel network and (b) baseflow, consisting of slow interflow and ground water flow. This permits the separation of stream flow into two, rather than three, components, simplifying the process of modeling and computation.

3. Hillslope Hydrology Hillslope hydrology refers to the hydrologic process taking place on hillslopes. These processes are intrinsically related to stream flow generation. The question to be answered in this

regard is whether the referred path of runoff is on the surface, as overland flow, or through the subsurface, by subsurface storm flow or fast interflow.

Theories of stream flow generation from hillslopes range from Hortonian overland flow (surface flow) to throughflow (subsurface interflow). Hortonian flow is applicable to arid and semiarid regions with poorly vegetated slopes and relatively thin soil covers. Throughflow is applicable to humid and subhumid regions with heavily vegetated areas and thick permeable soil layers. Between these two extremes lies a gamut of theories that consider a mix of Hortonian overland flow and throughflow.

The partial area flow concept (Ponce, 1989) is a combination of throughflow in the upper hillslopes and overland flow in the lower hillslopes. Interest in the partial area concept arose from the recognition that stream flow estimates were improved by assuming that only a small portion of the watershed is able to contribute to stream flow during the hydrograph peak. The size of the partial areas is a function of storm depth, rainfall intensity, and antecedent moisture. Generally, partial areas may be expected to encompass between 5 and 20% of the total watershed.

Refinements of the partial area concept have led to the variable source area model of hillslope hydrology. These variable source areas are envisioned as comprising low-lying lands adjacent to streams and rivers and concentrated near watershed outlets. Variations in the extent of source areas are dictated by antecedent moisture conditions, soil-moisture storage capacity, and rainfall intensity. When the upper soil horizon is saturated, both throughflow and overland flow occur, with eventual exfiltration to stream banks.

The variable source area model is a dynamic version of the partial area concept. This dynamism is manifested in soil moisture and runoff changes occurring annually, seasonally, between storms, and during storms. The variable source areas are largely responsible for overland flow, whereas the remainder of the watershed acts primarily as a reservoir to provide baseflow and to maintain saturation of the source areas.

4. Runoff Concentration An important characteristic of surface runoff is its concentration property. To describe this property, assume that a storm falling on a given watershed produces a temporally and spatially uniform effective rainfall intensity. In such a case, surface runoff eventually concentrates at the watershed outlet, provided the effective rainfall duration is longer than or equal to the time of concentration. Runoff concentration implies that the flow rate at the outlet gradually increases until rainfall from the entire watershed has had time to travel to the outlet and is contributing to the flow at that point. At that time, the maximum, or equilibrium flow rate is reached, implying that the surface runoff has concentrated at the outlet. The time that it takes a parcel of water to travel from the farthest point in the divide to the watershed outlet is referred to as the time of concentration.

The equilibrium flow rate is calculated by multiplying the effective rainfall intensity by the watershed area:

$$Q = 2.78\, I_e A \tag{6.1}$$

in which Q is the equilibrium flow rate in l/s, I_e is the effective rainfall intensity in mm/hr, and A is the watershed area, in ha.

The process of runoff concentration can lead to three distinct types of watershed response assuming constant rainfall excess and insignificant infiltration. The first type occurs when the effective rainfall duration is equal to the time of concentration. In this case, runoff concentrates at the outlet, reaching its maximum rate after an elapsed time equal to the time of concentration. Since rainfall stops at this time, subsequent flows at the outlet are no longer concentrated because not all the watershed is contributing. Therefore, the flow gradually starts to recede back to zero, as shown in Fig. 6.1(A), where I_e is effective rainfall intensity (mm/hr), Q is the flow rate (l/s), Q_e is the equilibrium flow rate (l/s), $Qmax$ is the maximum flow rate (l/s), t is time (hr), t_c is concentration time (hr), and t_r is the duration of effective rainfall (hr).

The second type of watershed response occurs when the effective rainfall duration is longer than the time of concentration. In this case, the runoff concentrates at the outlet, reaching its maximum rate after an elapsed time equal to the time of concentration. Since rainfall continues, the whole watershed continues to contribute to flow at the outlet, and subsequent flows remain concentrated and equal to the equilibrium value. After rainfall stops, the flow gradually recedes back to zero, as shown in Fig. 6.1(B).

The third type of response occurs when the effective rainfall duration is shorter than the time of concentration. In this case, the flow at the outlet does not reach the equilibrium value. After rainfall stops, the flow recedes back to zero. The requirement that volume be conserved and recession time be equal to the time of concentration leads to the idealized flat-top response shown in Fig. 6.1(C).

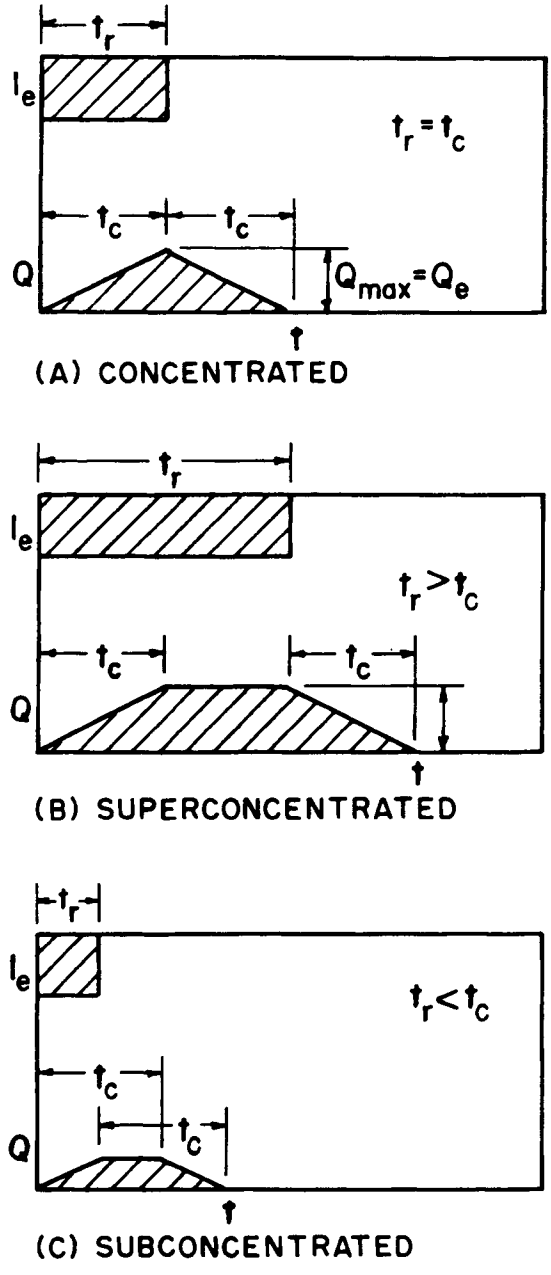


Figure 6.1—Effect of Rainfall Duration on Runoff.

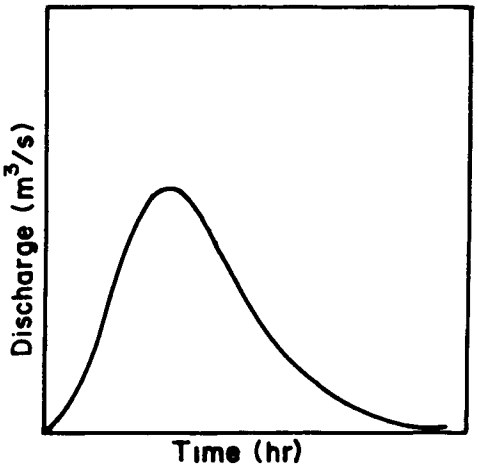


Figure 6.2—Typical Single-Storm Hydrograph.

5. Runoff Diffusion In nature, watershed response shows a more complex behavior than that which may be attributed solely to runoff concentration. Runoff rates from a watershed are governed by the natural processes of convection and diffusion. Convection refers to simple translation or wave travel. Diffusion is the mechanism acting to spread the flow rates in time and space.

The net effect of runoff diffusion is to reduce the flow rates to levels below those that could be attained by convection only. In practice, diffusion acts to smooth out watershed response. The resulting response function is usually continuous, and it is referred to as the stream flow hydrograph, runoff hydrograph, or simply "the hydrograph." Typical single-storm hydrographs have a shape similar to that shown in Fig. 6.2.

6. Rating Curves It is known that river stage varies with discharge, but the exact nature of this relationship is not readily apparent. Given a long and essentially prismatic channel reach, a single-valued relationship between stage and discharge at a cross-section defines the equilibrium rating curve. For steady flow, there is a single value of stage for each value of discharge. In this case, the equilibrium rating curve can be calculated with either the Manning or Chezy formulas. This property of uniqueness of the rating qualifies the channel reach as a channel control.

Rating curves are a useful and practical tool in hydrologic analysis, allowing the direct conversion of stage to discharge and vice versa. Discharge can be obtained from the rating curve by the simple procedure of measuring the stage. Conversely, if discharge is known, for instance, at a watershed outlet, stage at the outlet can be readily determined from a suitable rating.

There are several ways to determine an equation for the rating. Invariably, they are based on curve-fitting stage-discharge data. A widely used equation is the following:

$$Q = a(h - h_o)^b \tag{6.2}$$

in which Q is discharge (cms), h is the gage height (m), h_o is the reference height (m), and a and b are constants. Several values of reference height may be attempted. The proper value is that which makes the stage-discharge data plot as close as possible to a straight line on logarithmic paper, with the values of a and b determined by regression.

Other flow conditions, specifically nonuniform and unsteady, can cause deviations from the equilibrium rating. In particular, flood wave theory justifies the presence of a loop in the rating curve. Intuitively, the rising limb of the flood wave hydrograph has a steeper water surface slope than that of equilibrium flow, leading to greater flows at lower stages. Conversely, the receding limb has a milder water surface slope, resulting in smaller flows at higher stages (Fig. 6.3); however, the loop effect is likely to be small.

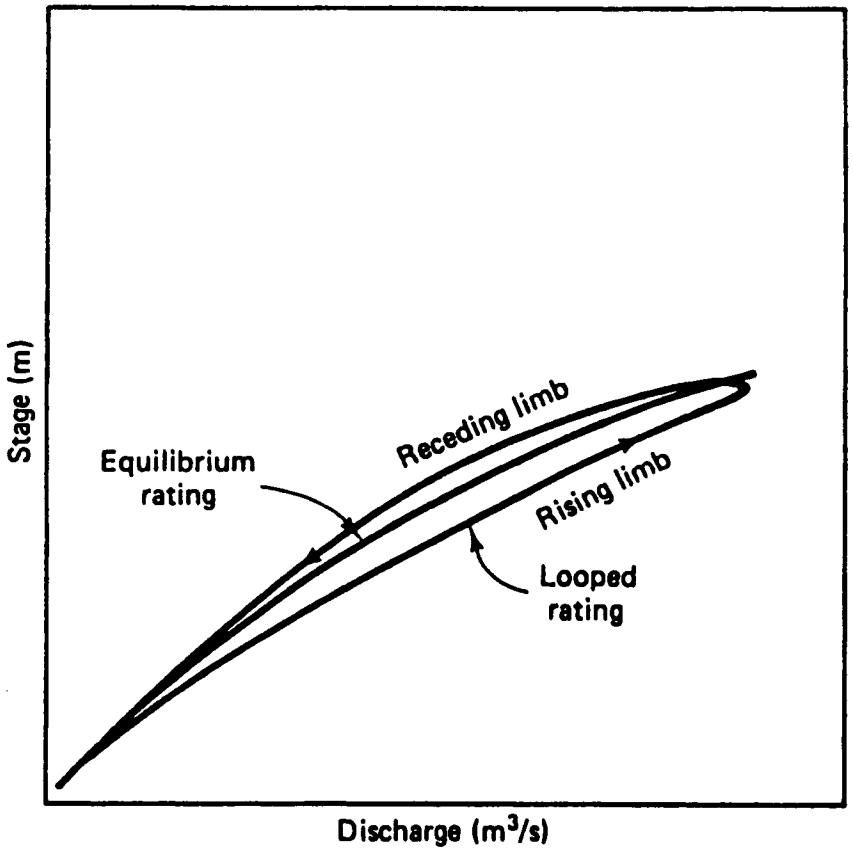


Figure 6.3—Equillibrium and Looped Rating Curves.

Two other mechanisms have a bearing in the evaluation of stage-discharge relations: the short-term and the long-term sedimentation effects. The short-term effects are due to the fact that the amount of boundary friction varies with flow rate. Rivers flowing on loose boundaries composed of gravel, sand, and silt constantly try to minimize their changes in stage. This is accomplished through the following mechanism: during low flow, the bottom friction consists not only of grain friction but also of form friction, the latter due to bed features such as ripples and dunes during high flows, the swiftness of the current acts to obliterate the bed features, reducing the form friction to a minimum, with only the grain friction remaining. The reduced friction during high flows gives rivers the capability to carry a greater discharge for a given stage. This explains the shift from low-flow rating to high-flow rating in natural river channels.

The long-term sedimentation effect is due to the fact that rivers continuously subject their channels to endless cycles of erosion and deposition, depending on the sediment load they transport. Some very active rivers may be eroding; others may be aggrading. Some geomorphologically active rivers may substantially change their cross sections during major floods. Invariably, shifts in rating are the net result of these natural geomorphic processes.

B. Variability of Runoff

Runoff and stream flow vary not only seasonally but also annually and with geographic location. Stream flow data for individual basins show that variability in runoff volume is significantly longer than that in the rainfall volume for the same period. On a global basis, total annual runoff volume is about 30% of the total precipitation volume. The difference is due to the abstractive processes, i.e., interception,

infiltration, surface storage, evaporation, and evapotranspiration. The value of 30% does not account for seasonal or geographical variability.

1. Seasonal Variability The typical basin in a temperate region shows runoff volumes varying during the year, low during the dry season and high during the wet season; however, basins in more extreme climates may show different behavior. For instance, in the ephemeral streams of typical arid regions, runoff is nonexistent during periods of no precipitation. For these streams, runoff and stream flow occur only in direct response to precipitation. Conversely, in humid and tropical climates, rivers show substantial quantities of runoff throughout the year, with comparatively little variability between the seasons.

The seasonal variability of stream flow can be traced to the relative contributions of direct and indirect runoff. During the wet season in temperate regions, indirect runoff is a small but nevertheless measureable fraction of direct runoff. In arid regions, particularly for ephemeral streams, indirect runoff is either negligible or does not exist. For humid regions, indirect runoff is substantial throughout the year and may be of the same order of magnitude as direct runoff.

The seasonal variability of runoff and stream flow can be further explained in the following way: ground water reservoirs serve as the mechanism for the storage of large quantities of moisture, which are slowly transported to lower elevations (or zones of lower potential). Some of the moisture is eventually released back to the surface waters. The process is slow and is therefore subject to substantial diffusion. The net effect is a permanent contribution from ground water to surface water in the form of dry-weather flow. Therefore, to evaluate the seasonal variability of runoff and stream flow, it is necessary to examine the relation between surface water and ground water.

2. Annual Variability Year-to-year stream flow variability shows some of the same features as those of seasonal stream flow variability. For instance, large basins show runoff variability from one year to the next as a function of the state of moisture at the end of the first year and of the precipitation quantities added during the second year. During dry years, rainfall replenishes the basin's stored moisture, with little of it showing as direct runoff. This results in the low levels of runoff that characterize dry years. Conversely, during wet years, the basin's storage capacity fills up quickly, and any additional precipitation is almost entirely converted into runoff, producing high stream flow levels that are characteristic of wet years.

Runoff and stream flow variability are intrinsically linked to the relative contribution of direct and indirect runoff involving the mechanisms of surface and ground water flow, and their temporal and spatial variability. The dearth of data for all relevant phases of the hydrologic cycle makes the evaluation of stream flow variability on mechanistic principles a rather complex process. A practical alternative is the reliance on statistical tools to develop estimates of runoff and stream flow variability based on records. This alternative has led to the concept of flood frequency, expressed as the average period of time (the return period) that it takes a certain flood magnitude to recur at a given location. An annual flood series is developed from daily discharge measurements at a given station. This is accomplished either by selecting the maximum daily flow for each of n years of record (the annual maxima series), selecting the greatest flow values in the entire n-year record (the annual maxima series), or selecting the greatest flow values in the entire n-year record, regardless of when they occurred (the annual exceedence series). The statistical analysis of the flood series allows the calculation of the flow rate associated with one or more selected frequencies.

3. Daily-Flow Analysis The variability of stream flow can also be expressed in terms of the day–to–day fluctuation of flow rates at a given station. Some streams show great variability from day to day, with high peaks and low valleys succeeding one another endlessly. Others show very little day–to–day variability, with high flows not very different from low flows. The reason for this difference in behavior can be attributed to differences in the nature of watershed response. Small basins are likely to have steep gradients and to concentrate flows with negligible runoff diffusion, producing hydrographs that show a large number of high peaks and correspondingly low valleys. Conversely, large basins are likely to have

milder gradients and to concentrate flows with substantial runoff diffusion, resulting in a smooth hydrograph with low peaks and comparatively high valleys. A practical way to evaluate day-to-day stream flow variability is the flow–duration curve (see section on flow duration curves).

4. Geographical Variability of Stream Flow Stream flow varies from one basin to another and from one geographic region of a certain climate to another of a given climate. Two variables help describe the geographical variability of stream flow: a) basin area and b) mean annual precipitation.

Intuitively, the volume available for runoff is directly proportional to the basin area; however, this concept is limited by the available precipitation. The mean annual precipitation determines the type of climate and therefore has a direct bearing on the variability of stream flow (see Chapter 2). The basin area is important, not only because of the potential runoff volume but also because larger basins have milder overall gradients. This causes increased runoff diffusion, increasing the chances for infiltration and loss of surface water to ground water. The net effect is that the peak discharge per unit basin area is inversely related to basin area, as in the following equation:

$$q_p = cA^{-m} \tag{6.3}$$

in which q_p is peak discharge per unit basin area, A is the basin area in km^2, and c and m are constants to be determined by regression. An example of this trend is represented by the Creager curves (Creager et al., 1945). The equation for Creager's enveloping equation is as follows:

$$q_p = C_1(0.386\text{A})^m \tag{6.4}$$

$$m = 0.9358\,A^{-0.048} \tag{6.5}$$

in which q_p is peak discharge per unit basin area, ((m^3/s)/km^2), A is the basin area (km^2), $C_1 = \text{C/A}$, and C is a constant varying in the range 38–130.

II. MEASUREMENT OF STREAM FLOW

The measurement of stream flow usually involves the determination of average velocity in the flow section and the application of this average velocity to an area of flow that is separately determined. This requirement for two types of data collection introduces several sources of possible error in the final determination of discharge rate. Many of the flow measurement techniques described herein deal primarily with the determination of average flow velocity. The errors quoted most often relate to velocity determination. The exceptions are flumes and weirs where velocity and area are combined to compute the discharge rate. When the area is defined by an irregular natural channel that is difficult to measure, significant errors in estimated stream flows may result due to errors or inaccuracies in the measurement or computation of flow areas.

A. Direct Measurement

This method involves measuring the width, depth, and velocity of flow in each of about 20 to 30 subsections of a stream cross section, computing the discharge in each subsection as the product of width, depth, and mean velocity, and summing the discharges in the subsections.

1. Current Meters The hydrographer selects the locations of the verticals at which velocity measurements are to be taken. The hatched area in Fig. 6.4 is a subsection whose vertical boundaries are halfway between the adjacent verticals. The verticals are located so as to obtain subsections which have about the same discharge, or ones for which the velocity in a subsection does not vary appreciably laterally from

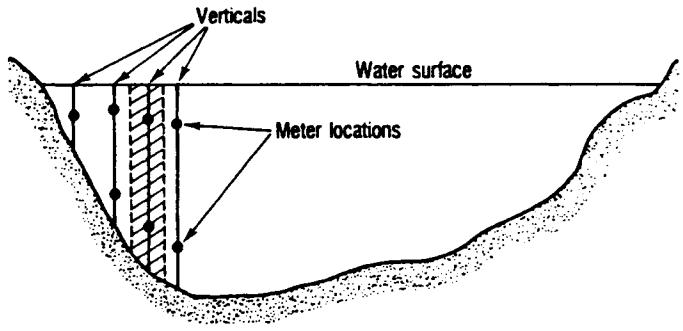


Figure 6.4—Stream Cross Section Showing Meter Locations for a Discharge Measurement (Riggs, 1985).

that measured. Fig. 6.4 also shows meter locations at which velocity measurements are made. These locations are at 0.2 and 0.8 of the depth except for shallow depths for which the meter is set at 0.6 depth from the water surface. Studies have shown the mean of velocities at 0.2 and 0.8 depth to be virtually the mean velocity in the vertical; similarly, the velocity at 0.6 depth is very close to the mean in the vertical. In special situations, the flow profile may warrant checking to verify that it is indeed "standard." This is done by collecting data at 10 or 20 depth points in one or more verticals and comparing this average velocity determination with the previously mentioned determinations (Riggs, 1985; Herschy, 1985).

Current meters commonly used in the United States are vertical-axis Price meters. The standard meter (Fig. 6.5) is for general use, and the Pygmy meter (two fifths the size of the standard meter) is used for shallow depths. Meters are rated by towing them at several constant velocities in a long tank containing still water.

When measuring velocity, the number of revolutions of the meter rotor is obtained by counting clicks (one click per one or five revolutions) in an earphone. Clicks are produced by an electrical circuit between the meter and the earphone. Elapsed time is measured with a stopwatch. The number of revolutions and the elapsed time can also be determined electronically with a current-meter digitizer, which also displays the results.

For measurements of shallow streams, the hydrographer usually wades directly into the stream. In the wading method, a tag line is stretched across the stream at the selected cross section. This line has marks

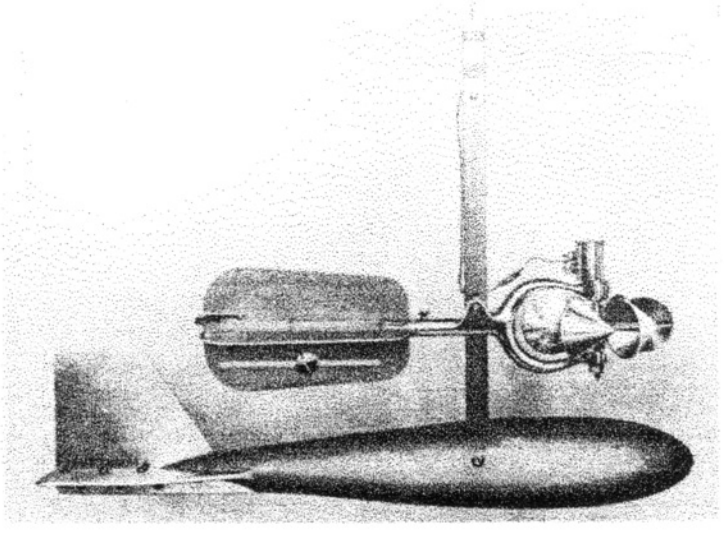


Figure 6.5—Standard Price Current Meter Mounted Above a Sounding Weight.

at selected intervals for horizontal control (usually 1 ft, 5 ft, 1 m, etc.). The current meter is mounted on a wading rod that is used to measure depth and to hold the current meter at the desired location or locations in the vertical.

Deeper streams are usually measured from bridges, cableways, or occasionally from boats. Bridge railings and cableways are marked at intervals for horizontal control. A heavy tag line is used for locating and holding the position of a boat on a cross section. The meter is suspended on a light cable above a sounding weight (Fig. 6.5). A reel mounted on the bridge, cableway or boat provides vertical location control. Flags at intervals on the cable and a counter on the reel are used to indicate the depth sounded and to set the meter to the desired positions in the verticals. The weight also helps maintain the line in a near vertical position.

When a stream is ice covered, measurements of depth and velocity may be made through holes cut in the ice along a selected cross section. The type of meter suspension depends on the water depth. Velocity distribution in the vertical under ice cover is similar to the velocity distribution in a pipe. Velocity is measured at 0.6 depth for depths under 0.75 m, and at 0.2 and 0.8 depth for greater depths. For greater precision, it may be desirable to thoroughly define the vertical velocity curves at one or more verticals in the section.

Reliability of discharge measurements ranges widely. Best results are obtainable in a smooth channel where the velocity is moderate, changing only slightly between adjacent verticals; the velocity is perpendicular to the cross section; and the discharge of the stream does not change during the measurement. Under these conditions, the potential accuracy of the measurement is rated excellent. Carter and Anderson (1963) discuss the accuracy of discharge measurements in more detail.

Portable electromagnetic velocity sensors are commercially available and are used in many of the flow metering situations that commonly use the mechanical meters discussed above. The sensor is a probe, 30 to 44 mm in diameter, which is pointed upstream. The basic procedure to determine channel discharge from these point velocity readings is the same as described for the mechanical meters. Water flowing past the probe interacts with an electromagnetic field generated inside the probe to produce a voltage in the water near the probe. This voltage is a function of water velocity. Auxiliary equipment facilitates data gathering and computation of the final discharge rate for a channel section. Its size allows it to be handled and used much like the Pygmy current meter except that it has selectable flow ranges that can generally allow it to also substitute for the larger mechanical meter for flows exceeding about 1 m/s. The necessary shielded cable that is used with it may limit its application from bridges.

These probes operate on electromagnetic principles. The probe consists of an electromagnet inside a molded plastic housing with two electrodes spaced 180° apart and about one probe radius from the rounded end of the probe. Water flowing around the probe end and along the probe interacts with the electromagnetic field to produce a small voltage in the water near the probe. This voltage is detected by the electrodes and is a function of the flow velocity. In one version, the result is processed and a point velocity is displayed on a panel meter.

With mechanical meters, the velocity errors are somewhat controlled by operator choices of count length and propeller selection. In the electromagnetic versions, these choices are averaging–time, corresponding to count length, and flow-range, which corresponds to propeller selection. Stray electrical currents, particularly when operating in urban environments, have been reported by field users.

Errors in area determination and velocity distribution frequently dominate the final discharge results when using the mechanical meters, particularly on small channels, meaning that the point velocities can be relatively accurate when compared to these other sources of error. The electromagnetic probes, on the other hand, may have problems of stable calibration and electrical interference that may introduce uncertainties in individual velocity determinations, adding another degree of uncertainty to the final results.

Because the reliability of velocity measurements obtained from the electromagnetic sensor is not comparable with the reliability of the mechanical meter, the International Standards Organization (ISO) has not yet approved this type of meter for use in gaging streams. A further limitation is the need for a

shielded cable between the probe and the recorder or indicator. This requirement appears to limit use of the electromagnetic velocity sensor to the gaging of small streams. The reader is referred to the section titled Electromagnetic Gaging Stations for an application of this technique.

2. ***Moving-Boat Method*** The employment of the moving-boat method permits the discharge of a large stream to be measured within a few minutes. The conventional velocity–area approach for determining discharge is used, but the data are collected at 30–40 verticals observed from a boat while it traverses the cross section.

A rate indicator unit is used in conjunction with a current meter rating table to obtain the vector combined velocity of the stream and the boat. An angle reading, representing the angle a vane makes with the cross-sectional path, defines the direction of the vector (Fig. 6.6).

Prior to measuring, two range markers on each side of the cross section are established, anchored floats are located on the cross section 15 to 20 m from each shore, and the width of the cross section is measured.

Also required are two or more vertical-velocity characterizations made in the cross section. Whereas the standard gage points may have one or two samples in each vertical, the vertical-velocity characterizations can be made with enough samplings at vertical intervals to reliably define the velocity profile in the vertical. This may require 10 to 20 point velocity determinations throughout the vertical. These are used to adjust velocities measured at a constant distance below water surface to mean velocities in the vertical.

Equipment on the boat includes an angle indicator, a propeller-type current meter and a sonic sounder. Also required is a rate indicator which shows the current-meter velocity and provides a method of automatically selecting measuring points at intervals of distance traveled.

The traverse of the cross section is made without stopping and data are collected at regular intervals. The boat operator maintains a course along the cross section by "crabbing" into the direction of flow sufficiently to keep on line. The force exerted on the current meter is the combination of two forces acting simultaneously, one force due to the movement of the boat, and the other from the stream flow normal to the cross section. Data collected at each sampling point include the sensed vector velocity of the boat and stream and the angle of the propeller and vane to the line across the stream, which permits computation of the stream velocity.

Details of the procedures and the computation of discharge are given by Smoot and Novak (1969) and Herschy (1985). Advantages of the moving-boat method over conventional methods include the short time required and the consequent reduction of interference with marine traffic. Reliability of results is comparable to that for conventional measurements, if the results of several traverses are averaged. In normal practice, a series of at least 6 runs, each with 30 or 40 verticals, are averaged.

Instruments currently used for the moving boat are more automated than those described above. Furthermore, an acoustic doppler current profiler that provides a vertical profile of horizontal velocities

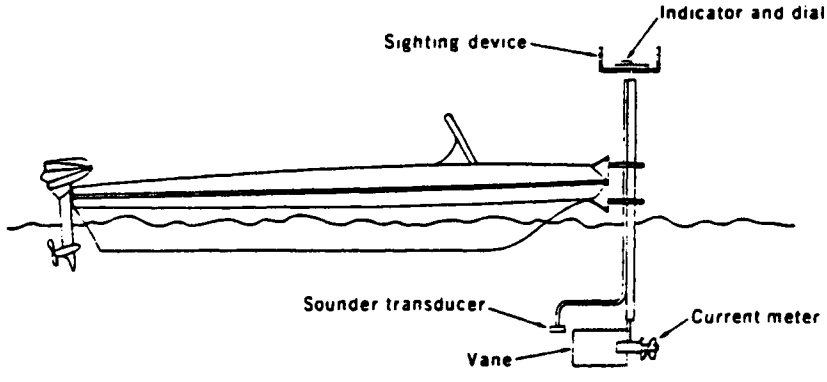


Figure 6.6—Equipment for Measuring by the Moving-Boat Method (Smoot and Novak, 1969).

and the direction of water movement is being developed. This system will likely develop as the new moving-boat system.

3. Other Methods Tracer-dilution methods are useful, primarily for small streams where there is considerable turbulence, making it difficult to use a current meter. These methods are also particularly useful when the cross-sectional area is difficult to define, such as boulder-filled mountain streams, ice-covered streams, or sand-channel streams. See Kilpatrick and Cobb (1985) for details.

Volumetric methods and portable weirs or flumes may be used to measure small flows of 500 l/s or less, but they are usually impractical in very rough channels or channels with very flat slopes. Permanently installed weirs or flumes usually produce reliable discharges once properly installed.

Unusual flood discharges have been estimated by timing drift in several parts of a cross section and by surveying the cross section after a flood. Results tend to be unreliable.

B. Indirect Measurements

It is frequently impossible to measure extreme flood peaks by the current-meter method because of inability to reach the site, the occurrence of floating or submerged debris, or the submergence or destruction of the measuring structure. Consequently, many peak discharges must be determined after the passage of a flood by indirect methods that are based on hydraulic equations relating the discharge to the water-surface profile and to the geometry and roughness of the channel. The water-surface profile is defined by high-water marks on both banks of the stream reach. See Benson and Dalrymple (1967) for the various types of high-water marks, and for general field and office procedures for making indirect measurements. The common types of indirect measurements are slope-area, contracted-opening, and flow-over-dams, weirs, and embankments.

1. Slope-Area This is the most commonly used form of indirect measurement. The basic equation in the United States is the Manning formula

$$Q = (1/n)\, A\, R^{2/3}\, S^{1/2} \tag{6.6}$$

in which Q is the discharge, (m^3/s), A is the cross sectional area (m^2), and R is the hydraulic radius (m). In United States customary units,

$$Q = (1.486/n)\, A\, R^{2/3}\, S^{1/2} \tag{6.7}$$

where Q is the discharge (cfs), A is the cross sectional area (sq ft), R is the hydraulic radius (ft), S is the friction slope (dimensionless), and n is a roughness coefficient.

A field survey of the selected reach involves locating and surveying high-water marks for defining water surface profiles on both banks, surveying three or more cross sections, and estimating the roughness coefficients for the reaches between adjacent cross sections. From these data, the discharge can be computed as described by Dalrymple and Benson (1967) and Herschy (1985).

The reach selected should be reasonably straight with uniform channel geometry, uniform channel roughness, a smooth water-surface profile, and no wide overflow sections. These conditions should also exist for some distance upstream of the reach. Length of the reach should be at least 75 times the mean depth in the channel. Fall in the reach should be greater than the velocity head and greater than 15 cm. A minimum of three cross sections are recommended.

Selection of the appropriate roughness coefficients requires some experience. Guidelines for selecting roughness coefficients in natural channels are given in many texts, for example, Chow (1959) and also Barnes (1967), who compiled the results of many verifications of the Manning roughness coefficient in natural channels. The roughness coefficient of a natural channel is a function of bed roughness, bank irregularity, effect of vegetation (if any), depth of water, channel slope, and other factors. Consequently,

the value of n computed from known water-surface profiles, channel geometry, and discharge may not be transferred with confidence to another site unless the dominant factors affecting n are common to the two sites.

Reliability of a discharge computed by the slope-area method is reduced if there are poor or questionable flood marks, if the water-surface profiles are irregular, and if the areas of adjacent cross sections are much different. Furthermore, the areas at cross sections, as measured after a flood, are assumed to be as they existed at the time of the peak discharge, but this is unlikely to be correct for channels with erodible beds. All of these factors, as well as the uncertainty in the estimates of n, need to be considered in evaluating the reliability of a particular slope-area measurement. Some indication of the reliability of peak discharges computed by the slope-area method is given by Riggs (1976) in which he uses estimates of the roughness coefficients recorded on the field notes of some surveys of channel reaches for which the coefficients were later verified from known peak discharges. These estimates of n (or averages of several) were available at 27 sites and were used to compute discharges from the data provided by the field surveys. The discharges, based on these estimates of n, were found to be within 30% of the known discharges. The estimates of n for these 27 sites were made by hydrologists with extensive experience in this type of work. Personnel with less experience might select less desirable reaches and make less suitable estimates of n.

2. Control in Channels Recording of stream flow usually depends on exploiting some type of channel control and recording a stream flow stage at a gaging station. This stage must be related to channel velocity and channel flow area. While area is a matter of channel geometry relative to stage, the flow velocity depends on a velocity control mechanism that can be mathematically characterized. For open channels, controls are frequently accomplished by channel control, which means that the flow velocity is controlled by channel friction. This rate of energy conversion to heat energy results in a decrease in total energy in the direction of flow (Chow, 1959). This slope of the energy line is used in equations such as the Manning formula to characterize flow in open channels. The errors of uncertainty in this control characterization can be considerable, 10 to 20%, or higher, because it usually requires an estimation of the friction value or current meter ratings to calibrate the channel section, which may not apply after storm flow passage. Because section calibrations usually change seasonally and during storm events, channel control is not as reliable as critical flow control.

Critical flow control, sometimes called section control, depends on an overfall such as a permanent riffle or rock ledge, a steep channel section, that causes supercritical flow downstream of a subcritical flowing reach, or a sufficient contracted opening, such as a bridge, that causes critical velocity. In this case, the contraction causes a rise in the upstream water surface that is just enough to pass the discharge through the opening at critical flow (Chow, 1959). Weirs and flumes are critical flow devices. The errors of uncertainty can be less than ±5% at well-characterized contractions or overfalls.

When suitable natural controls are not available, an artificial control may sometimes be attempted; but for large streams, artificial controls are expensive, likely to be damaged by floods and may cause changes in the approach channel.

3. Contracted–Opening Method Stream width contractions commonly occur at bridges. The contraction creates an abrupt drop in water-surface elevation between an approach section and the contracted section. This drop, the geometries of the approach and the contracted sections, and an estimate of the discharge coefficient are used to compute discharge. The equation is

$$Q = CA_3 \sqrt{2g\left(\Delta h + \alpha_1 \frac{V^2}{2g} - hf\right)} \tag{6.8}$$

in which Q is the discharge (m^3/s), C is the coefficient of discharge, A_3 is the gross area of contracted section (m^2), Δh is the difference in elevation between approach section and contracted section (m), α_1

$V^2/2g$ is the weighted average velocity head in the approach channel section, where V is the average velocity, Q/A, α_1 is a velocity distribution coefficient (Chow, 1959) that takes into account the variation in velocity in that section, and h_f is the head loss due to friction between the approach and the contracted sections (m).

The characteristics of a contracted section site should be carefully considered. A contraction should have a stable channel; the fall in water surface, as defined by high water marks, should be more than 15 cm; the fall should be more than four times the computed friction loss between the two sections; and the bridge geometry should be reasonably close to one of the types for which discharge coefficients have been defined. If these requirements cannot be met by any available contracted-opening site, some other method such as slope-area should be considered.

See Matthai (1967) for field and computation procedures including procedures for computing the discharge coefficient for selected bridge abutment geometries. Matthai also suggests methods for computing discharge through multiple-opening contractions and discharges that include flow over a road as well as through a contracted opening. The methods presented in Matthai (1967) are based on laboratory studies and field verifications.

4. Flow Through Culverts The peak discharge through culverts can be determined from high-water marks that define the headwater and tailwater elevations (Brater and King, 1976). For convenience in computation, culvert flow has been classified into six types based on the location of the control section and the relative heights of the headwater and tailwater elevations (Chow, 1959). Discharge equations have been developed for each type of flow by application of the continuity and energy equation between the approach section and the terminal section. Bodhaine (1968) describes the site information needed, how to determine the type of flow that occurred, and how to compute the discharge.

The peak discharge through a culvert is equivalent to the peak discharge of the stream a short distance above the culvert only if the flood caused minimal poundage upstream from the culvert. Therefore, poundage corrections are usually needed.

5. Flow Over Dams and Embankments A weir, dam, or embankment forms a control section at which discharge is related to the upstream water-surface elevation. The peak discharge at the control section can usually be determined from a field survey of high-water marks and the geometry of the particular structure.

The two most important elements of the flow computation are the head and the discharge coefficient of the structure; however, if a high velocity of approach makes the velocity head a substantial part of the total head, the computed discharge may be unreliable. Discharge coefficients have been derived for some types of dams and weirs. Thus, if the structure is one for which the discharge coefficient can be selected with confidence, and if the head on the structure can be accurately determined, then a reliable discharge can be computed.

Procedures for measuring peak discharge using dams, weirs, and embankments are given by Hulsing (1967). Also included in that report are discharge coefficients and formulas for three general classes of weirs—sharp-crested, broad-crested, and round-crested—and for highway embankments and weirs of unusual shape.

6. Weirs and Flumes More recently, a more complete treatment of fixed-section broad-crested weirs and long-throated flumes has become available (Bos et al., 1991). These configurations can be reliably rated by computer model (Clemmens et al., 1987) to within ±2% plus readout method error. Primarily developed for concrete-lined irrigation canals, the devices have been successfully used to measure discharges less than 25 m^3/s. Features are a) accuracy, b) economy of construction, c) flexibility of cross-sectional shape, d) fair to good sediment carrying capacity, and e) low required head loss, usually less than 10% of the stream flow depth. A typical configuration for the broad-crested weir style is shown in Fig. 6.7.

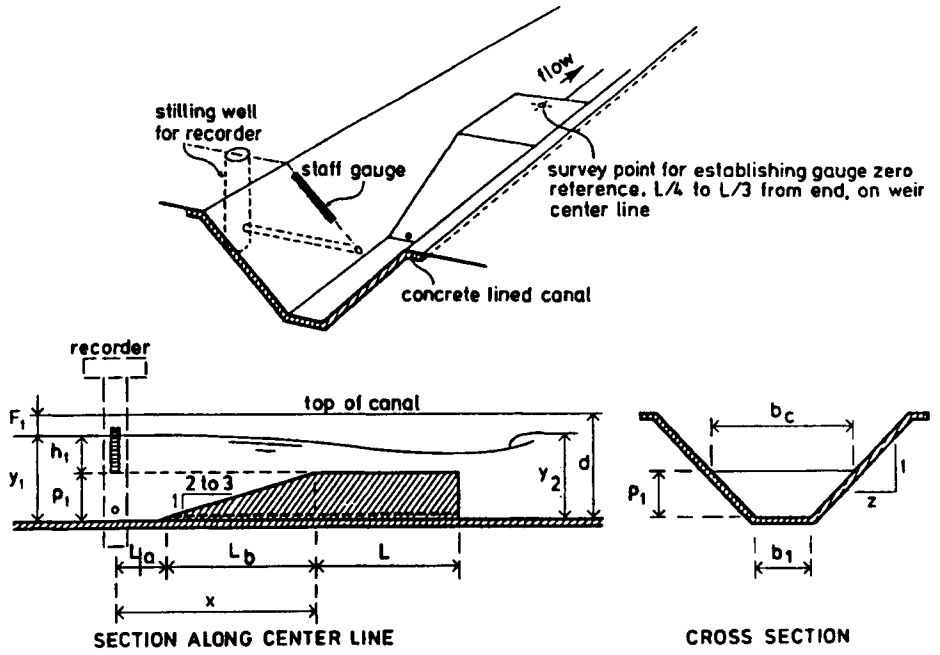


Figure 6.7—Typical Layout for Trapezoidal Long-Throated Flume (broad-crested weir style) Commonly Used in Lined Channels.

The broad-crested sill in a section of circular culvert or round pipe flowing partly full is shown in Fig. 6.8. The format shown is for a portable version that is practical for up to about 100 l/s.

Larger sizes have been designed for use in corrugated metal and concrete pipe culverts. Usually a sill height of one-quarter of the pipe diameter is used unless high downstream water levels force the use of a higher sill. The calibration for specific dimensions is derived by the modeling processes in Clemmens et al. (1987).

Sediments tend to clear these broad-crested weirs during the rising phase of a storm hydrograph, and appear to continue clearing during steady flow; however, on the falling leg of the hydrograph, the approach channel upstream from the sill may tend to fill with sediments. This is usually a problem with all weirs and flumes. Absolute velocities can be reliably computed for points throughout these devices, which can sometimes be used by the design engineer to evaluate sediment moving capabilities of known size or geomorphic source.

The computer calibration method is also applicable to side contractions. Complex sections for wide ranges in flow rate that approach 1:1000 (low to high flow rate ratio) also have been installed. In general,

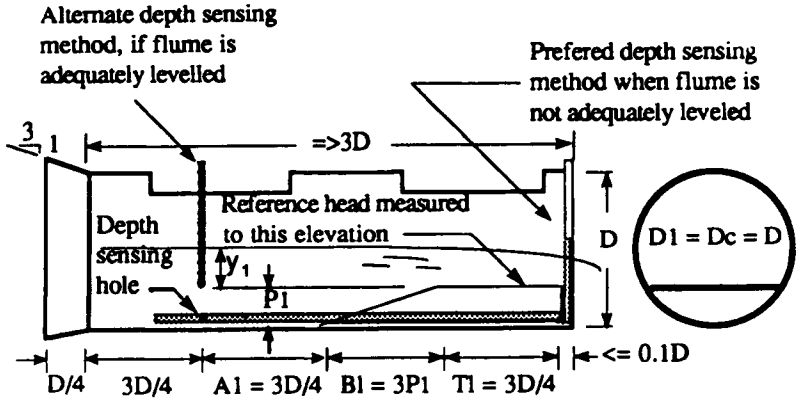


Figure 6.8—Broad-Crested Weir Style of Flume in a Round Pipe Flowing Partly Full.

the sediment carrying capacity of flumes does not appear to improve by using side contractions rather than bottom sills. A detailed treatment is given in Bos et al. (1991). Pre-computed tables are given for various rectangular and trapezoidal channels with sills in the previously mentioned references and in Clemmens et al. (1988) and Replogle et al. (1991). The latter also contains some precomputed sizes for circular pipes.

C. Continuous Records of Stream Flow

1. Conventional Gaging Stations Gaging stations typically obtain a continuous record of stream water stage. For channel control situations, this involves a continuous record of stream stage and the subsequent conversion to discharge through a rating curve. An example of a gaging station is shown in Fig. 6.9.

The general locations of gaging sites are usually dictated by water management considerations or by the hydrologic network. For channel control situations, an ideal location for a gaging station is a reach of stable channel that is isolated from downstream backwater effects by a steep section of channel, a permanent riffle, or a rock ledge a short distance downstream from the gaging station. At such a location, both reliable stage and calibration discharge measurements can usually be obtained. See Rantz (1982) for more detailed considerations in specific site selection.

2. Sensing and Recording Stage Water-surface elevation (stage) is usually referred to some arbitrary gage datum and is also called gage height. Gage height at a gaging station may be sensed by a float in a stilling well or by a gas-purge system that transmits the pressure head of water in the stream to a manometer or pressure transducer. Staff gages, mounted in the stream, are used to make periodic checks on the sensing equipment.

These sensings are recorded as a continuous trace of the gage height (on an analog recorder), as punches on a tape at intervals of a few minutes (on the digital recorders), or as computer bits on data acquisition systems. The gage-height record may also be recorded and stored temporarily on an electronic data collection platform and transmitted by satellite to computer data files. This and other applications of computer processing of hydraulic and hydrologic data is discussed in Herschy (1985).

3. Stage-Discharge Relations A stage-discharge relation is also called a rating or a rating curve. The relation is usually determined empirically by means of measurements of discharge and the corresponding stages. At a new station, many measurements are needed to define the relation throughout the entire range of stages. Periodic measurements are needed thereafter to verify the defined rating and identify changes, if needed.

When the discharge record at a new gaging station is needed before enough periodic data have been obtained to define the rating, the rating can be estimated using a survey of the downstream channel and

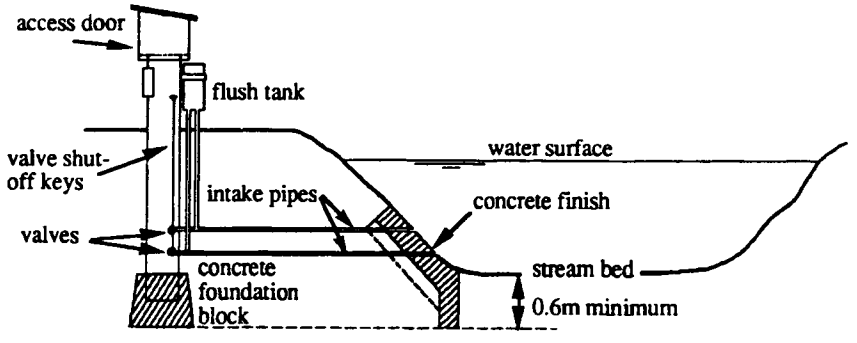


Figure 6.9—Example of an Intake Pipe System with Flush Tank that can be used at Stream Gaging Stations.

evaluations of channel roughness. The technique, known as the step-backwater method, is described in many hydraulics texts, e.g., Chow (1959), Davidian (1984), and verified by Bailey and Ray (1966).

Stage-discharge relations may change temporarily or permanently with changes in the physical characteristics of the channel. Consequently, as mentioned previously, periodic independent discharge measurements should be made to verify the established rating or to indicate changes. A temporary change in a rating may be caused by growth of vegetation in the channel, formation of ice or lodging of drift. Permanent changes usually result from scour or fill produced by a flood, or by some construction activity.

4. Discharge Relations Using Slope As Parameter No simple relation exists between stage and discharge in channels where water surface slope is variable. Variable slopes may be caused by a) backwater from a confluent stream or from tides, or by manipulation of gates at a downstream dam or b) by rapidly changing discharge in a channel having a relatively flat slope. The effect of the latter is to produce a loop curve on which the discharge for a given stage is greater when the stream is rising than it is when the stage is falling (Fig. 6.3).

Discharge ratings using stage and slope are of two types: constant-fall ratings and variable-fall ratings. See Rantz (1982), Kennedy (1984), and Herschy (1985) for procedures to develop these ratings. Constant-fall ratings are preferred where backwater is present at all stages and at all times (Chow, 1964; Scott, 1982; Herschy, 1985). Variable-fall ratings are used where backwater is intermittent.

To obtain the continuous record of slope needed for these ratings, an additional record of stream stage is collected at an auxiliary gage downstream from the base gage. Channel slope need not be measured where variable slope is caused by rapidly changing discharge. Methods are available to adjust the simple stage-discharge relation for the effect of changing discharge. Some of these methods are briefly discussed below. A detailed discussion is contained in Herschy (1985).

Many sites can be evaluated by stage-fall discharge ratings. In simple form, these consist of a reference gage at which the stage is observed continuously and current meter measurements are made occasionally, and a second gage downstream where stage is also observed continuously. Based on the same datum and the synchronized times of the two, the water surface fall in the distance between the two gages provides a measure of water surface slope.

Under backwater situations, the fall as measured between the two gages is used as a third parameter. It is variously referred to as the stage-fall-discharge method, or stage-slope-discharge method.

Stage-fall-discharge methods, as mentioned above, are broadly divided into a) constant fall ratings (with the unit-fall method as a special case), and b) variable-fall ratings.

The unit-fall-discharge method assumes that the ratio of Q/Q_r is proportional to the square root of the so-called fall ratio, F/f_c, where Q_r is the discharge corresponding to the stage at the first gage for a constant fall f_c, F is the measured water surface fall between the two gages, and Q is the current-meter discharge determination for that stage.

This square root assumption is most likely when the channel is approximately the same shape at both gage locations, if not the same size. The procedure is modified in the more general constant-fall rating procedure in that two curves are used. The first curve is the relation between stage and discharge for a constant fall of some specific value. The second is the relation between measured fall, F, and the discharge ratio, Q/Q_r. A unique feature of the constant fall method is that the two reference gages need not be at the same datum (Herschy, 1985).

Among the variable-fall ratings, the limited-fall rating is particularly applicable for situations when backwater is not always present. The general procedure is to deduce nonbackwater ratings based on measurements when backwater is observed. Later, the reverse procedure is used to obtain discharge rates that include corrections for the observed backwater. Because only stages are observed, a separate means is needed to detect backwater. This experience is usually obtained during the development of the initial ratings. When the backwater is variable, local experience is probably best in determining when to apply the backwater corrections.

These methods are subject to high levels of uncertainty and in complex cases can yield only approximate results.

D. Gaging Stations Using Index Velocity

1. Ultrasonic Meter Stations An ultrasonic velocity meter (UVM) measures the velocity of flowing water by means of an ultrasonic signal that moves faster downstream than upstream.

The major components of most UVM's are two transducers, one on each side of the channel, installed at the same elevation with the line between the two transducers diagonal to the stream, and a console on the streambank. The console provides the pulses to the system, records the responses to the pulses, and computes and records the index velocities with their corresponding times of occurrence.

These index velocities and the corresponding stages are calibrated with current meter measurements where possible.

UVM stations are used on large streams where variable backwater occurs and where stream slopes are too flat to provide discharge measurements accurate enough for defining an acceptable stage-fall-discharge relation. Theory and application are described by Laenen (1985).

2. Vane Meter Stations When flow in a small tidal stream may be in either direction and velocities are low, a vane (or deflection) meter may be used as an index of velocity. At such a site, the stage and angle of the vane are recorded continuously. Discharge measurements are used for calibration.

Vanes have been proposed for indicating flow rate per unit width based on measuring the horizontal force exerted on a triangular blade that is held rigidly, thus preventing deflection, which complicates the hydraulics. Restoring the blade to a vertical position with gravity weights by rotating the blade assembly about a horizontal axis is effective, but the blade shape is no longer triangular and is costly to fashion if the device is to indicate flow per unit width independent of flow depth. However, if a force cell of some sort is used in conjunction with a mounting about a vertical axis (door-hinge style) then only a simple triangular blade is required to indicate flow rate per unit width, independent of flow depth (Replogle, 1968).

3. Electromagnetic Gaging Stations This method permits the computation of a continuous discharge record on streams with highly unsteady flows and variable water-surface slopes. Continuous records of velocity at one point in the cross-section are obtained by an electromagnetic meter. A stage record at the cross-section is also obtained. Discharge measurements are used to relate discharge to stage and point velocity, stage being an index of cross-section area and point velocity an index of mean velocity in the cross-section. This relation is used to compute the continuous discharge record. Rantz (1982) describes the method in more detail.

One style uses an electromagnetic probe mounted on the floor of the channel that senses the near-wall velocity. Another sensor, usually in the form of a pressure transducer, senses flow depth. Previously developed data regarding the channel cross-sectional shape and the probable velocity profile allow a microprocessor unit to translate the two pieces of information into average section velocity and area for an instantaneous indication of discharge rate.

4. Discharge Records Downstream From Dams With Movable Gates Where other methods are impractical, the discharge through movable gates can be obtained by calibration with discharge measurements. Records of gate positions can then be used to compute a continuous record of discharge (Collins, 1977).

E. Partial Record Stations

1. Crest Stage Stations The purpose is to obtain, at little cost, a record of instantaneous flood peak discharges that are used to define a flood-frequency curve for the site. A crest stage gage is installed upstream from a high-water control. During a flood, the water rises in the gage pipe and carries at its

surface some powdered cork that is deposited on a stick in the pipe at the peak stage. After the flood, the hydrographer measures and records the height, erases the mark, and recharges the gage with cork. Discharges, corresponding to some of the peak stages, are defined by indirect discharge measurements. These discharge measurements are used to define a stage-discharge relation for the range of peak stages recorded. From this relation and the record of peak stage, a record of peak discharges is obtained.

2. *Low-Flow Partial Record Stations* Occasional measurements of low flow are made at a designated site over a period of two or more years. These measurements are correlated with the concurrent low flows of a nearby gaged stream for which a low-flow frequency curve has been defined. The object of operating the partial-record station, to obtain an estimate of the 7-day 10-year low flow at the site, is accomplished by transferring the 7-day 10-year low flow of the gaged stream through that correlation (Riggs, 1972).

An example of the procedure and an empirical test of the reliability of the estimates of the 7-day 10-year estimate at a partial-record station are reported by Riggs (1985). Estimates are quite reliable if the concurrent base flows of the two streams are closely related, if the needed downward extrapolation of the relation is small and if the estimate is larger than 30 l/s.

3. *Mean-Flow Sites* Mean flow can be estimated from monthly measurements made at a site for a year. The measurements should be made about the middle of each month and should be used with concurrent discharges and computed monthly mean flow at a nearby gaged stream. See Riggs (1985) for further elaboration and an example of the application.

III. HYDROGRAPHS

A hydrograph is a graphical representation of the variation of a water flow quantity (discharge or stage) with respect to time. Depending on the represented flow quantity, the hydrograph is designated as surface runoff, base flow, stream flow, low flow, flood flow, or stage hydrograph. The runoff or discharge hydrograph at a point represents the response of the contributing watershed to rainfall and/or snow melt, antecedent precipitation, infiltration, evaporation, evapotranspiration, and topography. Depending on the time scale used, one can develop a continuous stream flow hydrograph for several successive years, an annual stream flow hydrograph or an event-based flood or storm flow hydrograph. Continuous stream flow hydrographs are used for reservoir operation studies and power generation studies for run-of-the-river hydroelectric power plants. The event-based hydrographs result from relatively short-duration events rather than long-term (e.g., a month or a year long) weather patterns. Event-based hydrographs are used to develop design flood hydrographs for spillways and floodways, e.g., probable maximum flood, standard project flood, and 100-year flood hydrographs. These hydrographs are defined by the storm that produces them, e.g., 6-, 24-, 48-, or 96-hour probable maximum precipitation, standard project storm, and 100-year precipitation. This section describes methods to develop event-based or storm flow hydrographs and their components.

A. Hydrograph Components

A typical storm runoff hydrograph is shown in Fig. 6.10. It consists of three distinct components: the approach limb AB, the rising limb BD, and the receding limb DH. Addition of a fourth component, i.e., baseflow, to the storm runoff hydrograph results in the complete stream flow hydrograph. The shape of the approach limb depends on the antecedent conditions, distribution of rainfall intensity, hydraulic length of basin, ground slope, and hydrologic conditions of soils in the lower portions of the drainage basin. The rising limb depends on the same characteristics along with the duration and temporal and spatial distribution of rainfall in the entire watershed. The receding limb begins with the decrease in the amount of or the cessation of surface flow to the channel and represents discharge due to drainage of water stored in the channels and floodplains. The receding limb, particularly in the lower part, also

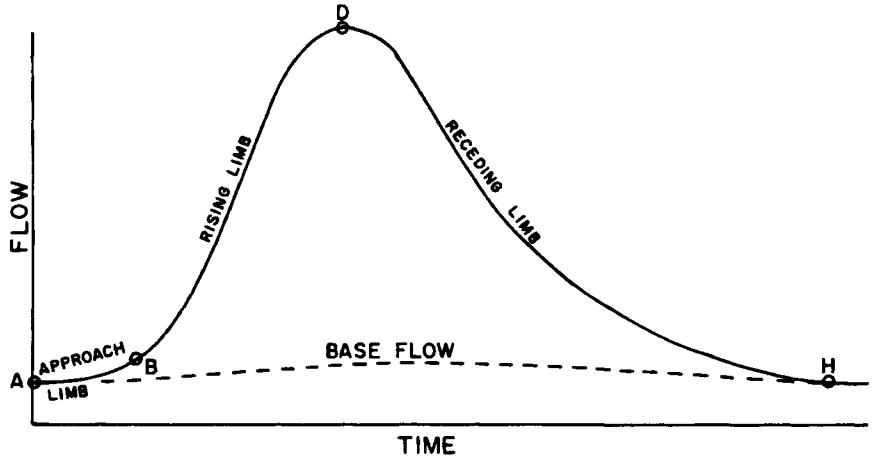


Figure 6.10—Typical Runoff Hydrograph.

includes withdrawals from bank storage and interflow (lateral flow) through surface soils. The lower portion of the receding limb exponentially decreases to the baseflow of the stream that is contributed only by the discharge of ground water.

The approach and rising limbs and upper portion of the receding limb of a surface runoff hydrograph are generally estimated by the unit hydrograph procedure. The recession limb is estimated by empirical and heuristic methods or analytical and numerical ground water flow models (see Fig. 6.11).

B. Drainage Basin Effects

The hydraulic characteristics of a watershed have two basic effects on the shape of the hydrograph at the outlet of the watershed. The first effect is due to the size, shape, hydraulic length (length measured along the average flow path of surface runoff), slope, and drainage pattern of the watershed. This effect is reflected in the unit hydrograph of the basin and includes the travel times of overland and channel flows and detention and storage of runoff in surface features, e.g., lakes, swamps, and drainage channels. The second effect is due to infiltration and detention in surface soils and other superficial features, interception and evapotranspiration by plants, and evaporation from soil surfaces. This effect is reflected in precipitation losses and transforms total precipitation to precipitation excess or effective precipitation.

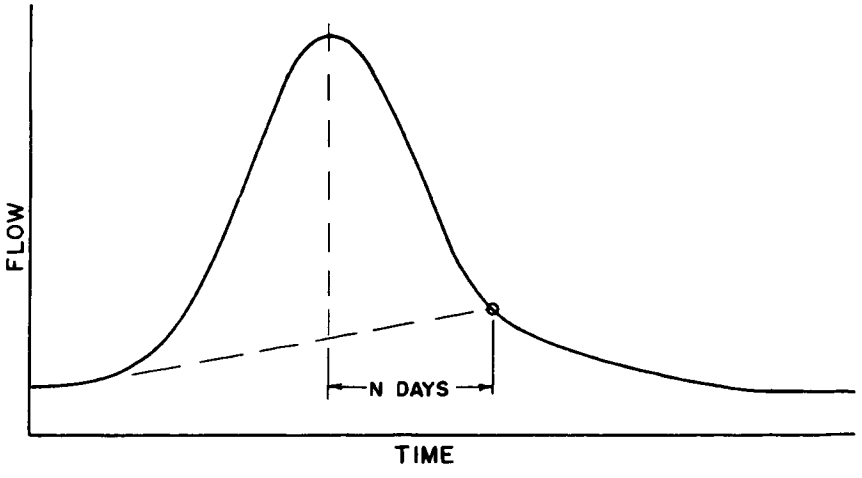


Figure 6.11—Empirical Truncation of Recession Limb.

C. Estimation of Precipitation Losses

Precipitation losses include initial abstraction, infiltration, interception, evapotranspiration, and depression storage. Methods to estimate infiltration loss include field measurements using infiltrometers, computation of average rate of infiltration (W - index or Φ index) using time-intensity graphs of rainfall and observed volumes of the corresponding storm runoff, and use of calibrated theoretical models (e.g., Philip, Horton, and Green and Ampt models; see Chapter 3). Methods to estimate evapotranspiration loss include field measurements, use of relevant evapotranspiration data and use of theoretical evapotranspiration models (Chow, 1964; Ponce, 1989; Eagleson, 1970).

There are four commonly used methods to account for precipitation losses in a rainfall-runoff analysis (Barfield et al., 1981; USACE, 1987; Ponce, 1989).

1) Constant initial loss followed by uniform loss rate. In this method, the entire precipitation is lost until the prescribed initial loss is satisfied. Thereafter, precipitation loss occurs at the prescribed uniform rate. This method is suitable for watersheds where calibrated or experimentally verified data for initial and uniform loss rates are available for various antecedent soil moisture and climatic conditions.
2) Empirical Loss Rate Function. In this method, the loss rate, L_R, in mm or inches per hour

$$L_R = (A + D)\, P^e \tag{6.9}$$

in which P is precipitation depth in mm or inches during the selected time interval, e is the exponent relating precipitation to loss rate, usually varying from 0 to 1.0, and A and D are variable loss rate coefficients, such that,

$$A = S/R^{0.1C} \tag{6.10}$$

and

$$D = 0.2\, D_R \left(1 - \frac{C}{D_R}\right)^2. \tag{6.11}$$

S is the starting or maximum value of the exponentially varying component of the loss rate coefficient (A), usually a function of the infiltration capacity or soil type, land use, and vegetal cover; R is the ratio of loss rate coefficient on the exponential loss rate curve to that corresponding to 25.4 cm (10 inches) more of accumulated loss, usually dependent on the capacity of the surface soils of the basin to absorb precipitation; C is accumulated rain loss up to the current time interval; D_R is the parameter defining the component D of the variable loss rate coefficient and its variation with cumulative rain loss; and D_R is primarily a function of the antecedent soil moisture deficiency and is usually storm-dependent.

Because of dimensional incongruity, there is no direct conversion for the coefficients of this method between the English and metric units. In each case, the coefficients have to be estimated by separate calibrations. Physical visualization of the values of the coefficients of this method is rather difficult. Therefore, its application is limited to watersheds for which calibrated values are known *a priori* or reliable coefficients can be obtained by parameter optimization.
3) Soil Conservation Service (SCS) Curve Number Method (USDA, 1985). In this method, a curve number (CN) is assigned to each watershed or portion of watershed based on soil type, land use and treatment, and antecedent moisture condition. The SCS has classified more than 4,000 soils into four hydrologic soil groups. Local SCS offices can provide pertinent information on soils in their respective jurisdictions. The definitions of the four hydrologic soil groups are reproduced below:

a) Soils having high infiltration rates even when thoroughly wetted and consisting chiefly of deep, well- to excessively-drained sands or gravels. These soils have a high rate of water transmission and a low runoff potential.
b) Soils having moderate infiltration rates when thoroughly wetted and consisting chiefly of moderately deep to deep, moderately well- to well-drained soils with moderately fine to moderately coarse textures. These soils have a moderate rate of water transmission.
c) Soils having slow infiltration rates when thoroughly wetted and consisting chiefly of soils with a layer that impedes downward movement of water, or soils with moderately fine to fine texture. These soils have a slow rate of water transmission.
d) Soils having very slow infiltration rates when thoroughly wetted and consisting chiefly of clay soils with a high swelling potential, soils with a permanent high water table, soils with a claypan or clay layer at or near the surface, and shallow soils over nearly impervious material. These soils have a very slow rate of water transmission and a high runoff potential.

The moisture condition resulting from precipitation 5 to 30 days preceding the storm under consideration is referred to as the antecedent moisture condition (AMC). Antecedent moisture conditions are divided into the following three groups;

- AMC-I—A condition of watershed soils where the soils are dry but not to the wilting point, and when satisfactory plowing or cultivation takes place. This condition is not considered applicable to design basis flood computations.
- AMC-II—The average case for annual floods, that is, an average of the conditions which have preceded the occurrence of the maximum annual flood on numerous watersheds.
- AMC-III—When heavy rainfall or light rainfall and low temperatures have occurred during the 5 days previous to the given storm, and the soil is nearly saturated.

Curve numbers (CNs) for various soils and land uses for the average watershed condition, AMC II, are given in Chapter 3. Curve numbers for one antecedent moisture condition may be converted to a different one by the use of Table 6.1. Alternatively, curve numbers for AMC I (CN_I) and AMC III (CN_{III}) can be computed from those for AMC II (CN_{II}) using the following equations (Hawkins et al., 1985):

$$CN_I = \frac{CN_{II}}{2.3 - 0.013\ CN_{II}} \tag{6.12}$$

$$CN_{III} = \frac{CN_{II}}{0.43 + 0.0057\ CN_{II}}. \tag{6.13}$$

It should be noted that the curve number method is an empirical method to estimate surface runoff. Precipitation loss is estimated as the difference between precipitation and computed surface runoff. The equations used to estimate surface runoff or rainfall excess are as follows:

$$Q = \frac{(P - 0.2S)^2}{P + 0.8S}, \quad P > 0.2S \tag{6.14}$$

$$S = \frac{2540}{CN} - 25.4 \tag{6.15}$$

in which Q is surface runoff (*cm*), S is the potential soil retention (*cm*), and P is precipitation (*cm*). The curve numbers have been found to decrease with increasing rainfall. The precipitation depth, P, and curve number, CN, used in Equations 6.14 and 6.15 are independent of storm duration, and

TABLE 6.1. Corresponding Runoff Curve Numbers for Three AMC Conditions.

AMCII	AMCI	AMCIII	AMCII	AMC1	AMCIII
100	100	100	60	40	78
99	97	100	59	39	77
98	94	99	58	38	76
97	91	99	57	37	75
96	89	99	56	36	75
95	87	98	55	35	74
94	85	98	54	34	73
93	83	98	53	33	72
92	81	97	52	32	71
91	80	97	51	31	70
90	78	96	50	31	70
89	76	96	49	30	69
88	75	95	48	29	68
87	73	95	47	28	67
86	72	94	46	27	66
85	70	94	45	26	65
84	68	93	44	25	64
83	67	93	43	25	63
82	66	92	42	24	62
81	64	92	41	23	61
80	63	91	40	22	60
79	62	91	39	21	59
78	60	90	38	21	58
77	59	89	37	20	57
76	58	89	36	19	56
75	57	88	35	18	55
74	55	88	34	18	54
73	54	87	33	17	53
72	53	86	32	16	52
71	52	86	31	16	51
70	51	85	30	15	50
69	50	84			
68	48	84	25	12	43
67	47	83	20	9	37
66	46	82	15	6	30
65	45	82	10	4	22
64	44	81	5	2	13
63	43	80	0	0	0
62	42	79			
61	41	78			

Source: USDA 1985.

the resulting precipitation loss includes initial abstraction, infiltration, interception, evaporation, and surface detention.

4) Holton Loss Rate Method. This method uses the following equation for the rate of infiltration in mm or inches per hour, f,

$$f = G \cdot A \cdot F^b + f_c \tag{6.16}$$

in which G is a growth index representing the relative maturity of the ground cover varying from near 0.0 when the crop is planted to 1.0 when the crop is full-grown; A is the infiltration capacity in mm or inches per hour that is related to the cover crop and soil texture and is interpreted as an index of the pore volume directly connected to the soil surface; F is the equivalent depth of unfilled pore space in mm or inches in the surface soil layer that is available for storage of infiltrated water, ranging from maximum available soil moisture capacity to 0.0; f_c is the constant rate of percolation

below the surface layer in mm or inches per hour; and b is an exponent with a typical value of 1.4. Holton et al. (1975) have given the values of A for several vegetation types. The values of f_c range from 11.4 to 7.6 mm/hr (0.45 to 0.30 in/hr) for hydrologic soil group A, 7.6 to 3.8 mm/hr (0.30 to 0.15 in/hr) for B, 3.8 to 1.27 mm/hr (0.15 to 0.05 in/hr) for C, and 1.27 mm/hr (0.05 in/hr) or less for D. The quantity of infiltrated water during a time interval is taken to be the quantity of available water (i.e., rain or snowmelt), or the estimated loss in mm (inches), F, during the time interval, whichever is smaller. The quantity, F, is computed by an iterative process as follows:

$$F = \frac{f_1 + f_2}{2} \Delta t \tag{6.17}$$

f_1 is the infiltration rate in mm/hr (inches/hour) at the beginning of the time interval computed from Equation 6.16 with known depth of unfilled pore space at that time (i.e., S_1), and f_2 is the infiltration rate in mm/hr (inches/hour) at the end of the time interval computed with an assumed value of unfilled pore space at the end of the time interval (i.e., S_2). The value S_2 is successively refined using the following relationship,

$$S_2 = S_1 - F + f_c \Delta t. \tag{6.18}$$

This method has a tractable physical basis; however, a number of soil and cover parameters must be known from previous studies, experiments, or calibration to use the method for a practical situation. In addition to the ranges indicated previously, default values of the parameters of the method are given in the HEC-1 Users Manual (USACE, 1987). Using these values as initial estimates, the unit graph and loss rate optimization option of the HEC-1 model can be used to obtain values that will best reconstitute the observed runoff hydrograph for a sub-basin.

D. Hydrograph Recession and Baseflow Separation

The upper portion of the recession limb of a hydrograph is contributed by 1) the portion of infiltrated water that passes through the shallower zones of the soil to reach the stream within a relatively short time after the storm without reaching the ground water table (i.e., interflow) and 2) exfiltration from bank storage of water that infiltrates into the river banks during high river stages. The portion of the recession limb from the inflection point to the point of return to baseflow usually plots as a straight line or a series of straight lines on semi-logarithmic paper, with flow plotted on the log scale. This leads to the following forms of recession equations:

$$Q(t) = Q_o\, e^{-t/ts} \tag{6.19}$$

or

$$Q(t) = Q_o\, K^t \tag{6.20}$$

in which $Q(t)$ is the flow at any time t after the inflection point on the recession limb, Q_0 is the flow at the inflection point, t_s is a recession constant known as the time of storage, and K is a recession constant. Values of the recession constants, t_s and K, are obtained by analysis of observed storm runoff hydrographs.

Baseflow is the dry weather flow of perennial streams contributed by subsurface flow from the water table due to infiltration antecedent to the runoff period under study, supplemented by any

recharge resulting from deep percolation of infiltrated water during the period under study. In most cases, baseflow constitutes a small portion of the total storm runoff. Therefore, minor errors in its estimation do not significantly affect design-basis flood or storm runoff hydrographs; however, it is necessary to know the baseflow hydrograph to determine surface runoff and unit hydrographs from a given stream flow hydrograph and to develop a complete flood hydrograph using the unit hydrograph for a basin. Some commonly used methods for baseflow separation are described in the following paragraphs.

1) Empirical Truncation of Recession Limb (Linsley et al., 1958). In this method, the recession limb of the storm runoff hydrograph is truncated N days after the time of peak flow, where,

$$N = 1.21\,A^{0.2} \quad \textit{where } A \textit{ is in } km^2$$

$$N = A^{0.2} \quad \textit{where} A \textit{ is in } mi^2 \tag{6.21}$$

in which A is the drainage area. The values of N given by Equation 6.21 should be slightly reduced for mountainous watersheds and increased for long narrow or relatively flat basins. This point on the recession limb is connected backwards to the beginning of the approach limb by a straight line (Fig. 6.12).

2) Extension of Curve Preceding the Approach Limb. In this case, the portion of the hydrograph preceding the approach limb is smoothly extended to a point in time under the peak of the storm runoff hydrograph. From this point, a straight line is drawn to a point N days after the peak on the recession limb. This implies that there is subsurface flow from the stream to the banks as the water surface in the stream rises resulting in gradual decrease in baseflow during the rising time of the hydrograph (Fig. 6.12).
3) Backward Extension of Lower Portion of Recession Limb. In this method, the lowest portion of the recession limb is smoothly extended backwards to the time of inflection on the recession limb. This point is joined to the beginning of the approach limb by an arbitrarily drawn rising limb. This approach represents situations where ground water contribution is relatively large and reaches the stream fairly rapidly (Fig. 6.13).
4) Use of Semilogarithmic Plots of Hydrographs (Dooge, 1973). When a storm flow hydrograph is plotted on semi logarithmic paper with flows on the log scale, the approach limb and upper and lower portions of the recession limb may be approximated by separate straight lines. The lower

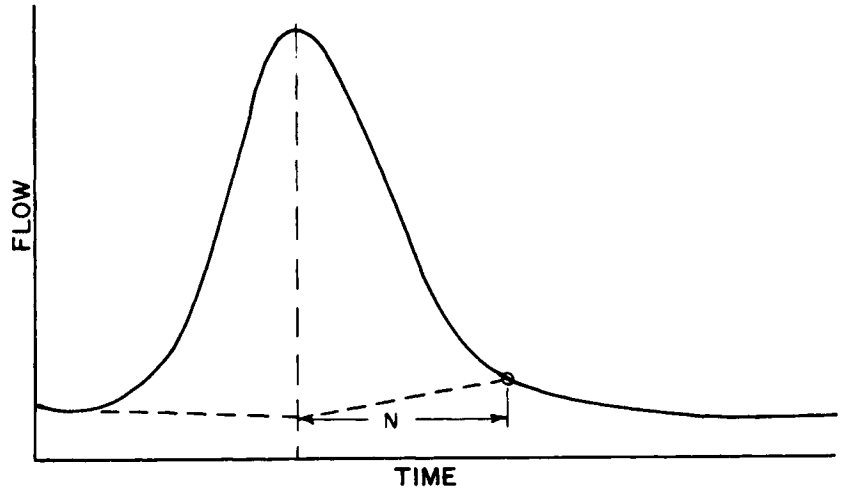


Figure 6.12—Extension of Curve Preceding Approach Limb.

Chapter 7

SNOW AND SNOWMELT

I. INTRODUCTION

This chapter considers snow as a factor in the hydrologic cycle and, in turn, in engineering applications. Involved are such questions as: 1) whether or not precipitation falls in a liquid or solid state; 2) how much snow (more specifically, "snow water equivalent") is present at the beginning of a snowmelt event; and 3) how rapidly snow on the ground will melt when meteorological conditions are favorable. The hydrologic analysis required to address such questions can include consideration of atmospheric energy fluxes, the seasonal water balance of the watershed, and the effect of the snowpack on the storage and delay of liquid water derived from either snowmelt or rainfall. Techniques are introduced in this chapter to cover such considerations in a practical manner, with options to utilize procedures of varying degrees of sophistication. The techniques described herein are meant to be applied to various types of studies and uses, including: 1) design flood determinations for planning and design; 2) day-to-day and seasonal runoff forecasting of river flows and water levels; and 3) hydrologic studies of a general nature for defining the effect of snow on runoff.

The techniques of analyzing snow in the hydrologic cycle and for quantifying snowmelt runoff rates are described generally in terms of computer modeling of some sort, whether it be a single-event analysis of flood runoff or continuous simulation throughout the accumulation and ablation of the snow. In addition, hydrologic regimes subject to snow lend themselves to statistical modeling principally for seasonal runoff volume forecasting, which is also described. The application of snow hydrology generally involves two distinct types of runoff situations: 1) rain-on-snow events typical of the winter floods in the Sierra Nevada and Cascade Mountains of the western United States and the Appalachians in the East; and 2) snowmelt-dominated runoff that is typified by spring/summer floods on large rivers such as the Columbia, Missouri, and Colorado.

The scope of this chapter is intended to provide only relatively brief descriptions of practical applications in snow hydrology for practicing engineers and hydrologists who do not have detailed experience in this subject. Extensive descriptive coverage of the physical processes involved and in-depth explanations of theoretical concepts are not included. Since the proper application of principles contained herein requires considerable background, the use of reference material cited in this chapter is highly recommended.

II. OVERVIEW OF PHYSICAL PROCESSES

A. Precipitation, Snowfall, and Snow Accumulation

In the middle latitudes, precipitation usually occurs as a result of the colloidal instability of a mixed water–ice cloud at temperatures below 0°C. The formation of snow and rain in the atmosphere is a

dynamic process. It has been observed that winter precipitation occurs initially in the form of snow crystals in sub-freezing portions of clouds. As the snowflakes fall through the atmosphere, they later melt into raindrops when they fall through warmer, above-freezing air at lower elevations. The melting level air temperature for snowflakes falling through the atmosphere varies from 0–4°C, but it is usually about 1–2°C. Accordingly, on the earth's surface, snowfall occurs at elevations higher than the melting level, while rainfall occurs at elevations lower than the melting level. The most significant determinant of the occurrence of rain or snow is the elevation of the melting level. This is particularly important in mountainous regions. Factors influencing the amount and distribution of precipitation in the form of snow and the snowpack water equivalent may be classified as being meteorologic and topographic. Meteorologic factors include air temperature, wind, precipitable water, atmospheric circulation patterns, frontal activity, lapse rate, and stability of the air mass. Topographic factors include elevation, slope, aspect, exposure, and vegetation cover.

B. Snow Metamorphism

Freshly fallen dry snow exists in a clearly defined crystalline state, with sharply defined edges and abrupt points in each snow crystal. Metamorphism of the snow occurs over time as the individual crystals lose their original distinct form and become rounded and bound together, ultimately into uniform, coarse, large ice crystals. This transformation may take place in a period of time as short as several hours but commonly involves a period of days, weeks, or months in intercontinental areas with large, deep snowpacks. Accompanying the change in crystalline character is a general increase in the density of the snowpack and a change in the temperature of the snowpack. This entire process, commonly called "ripening," is brought about by heat exchanges at the snow surface and ground, the percolation of rain and melt through the snowpack, and snowpack settlement and wind packing. Fig. 7.1 graphically illustrates a typical seasonal representation of snowpack condition as measured at the Central Sierra Snow Laboratory (USACE, 1956). This chart shows depth, density, and temperature of the snow throughout the season, together with positions of the layers of the snowpack as determined by a special settling gage.

Snow density is generally expressed as the ratio (in percent) of the weight of a unit snow volume to the weight of a unit volume of liquid water. The density of the newly fallen snow is typically on the order of 10%, with variations of 6–30% dependent upon the meteorological conditions involved (primarily air temperature and wind). As metamorphism occurs, density increases, reaching values of 45–50% at the end of the season. A snowpack ripe for melt also contains a small amount of free water, on the order of 3–5%. A ripe snowpack is said to be "primed" to produce runoff when it contains all the water it can hold against gravity. The snowpack temperature usually varies seasonally throughout the winter to spring period. In its early stages, the variation throughout the depth may be marked, from near 0°C at the ground to sub-freezing temperatures at shallower depths. As the snow reaches its primed condition, the snowpack is completely isothermal and near 0°C.

The snow condition is an important consideration in snow hydrology since a cold, dry snowpack can act as a temporary storage reservoir for meltwater or rain; however, it should be noted that the snowpack condition can change rapidly at any season of the year, as the result of heavy winter rain, accompanied by unusual winter snowmelt. In such situations, the snowpack may become fully primed in a matter of hours, with the pack isothermal and liquid water requirements fully met. The change in temperature gradient during the period from the fifth to the fifteenth of January in Fig. 7.1 is an illustration of such an occurrence. Under these conditions, subsequent rainfall and/or snowmelt will pass through the pack, regardless of the snowpack density, with little or no time delay.

C. Snowmelt

The process of melting snow involves the transformation of snow/ice from its solid form to liquid water through the application of heat energy from outside sources. While the latent heat of ice is

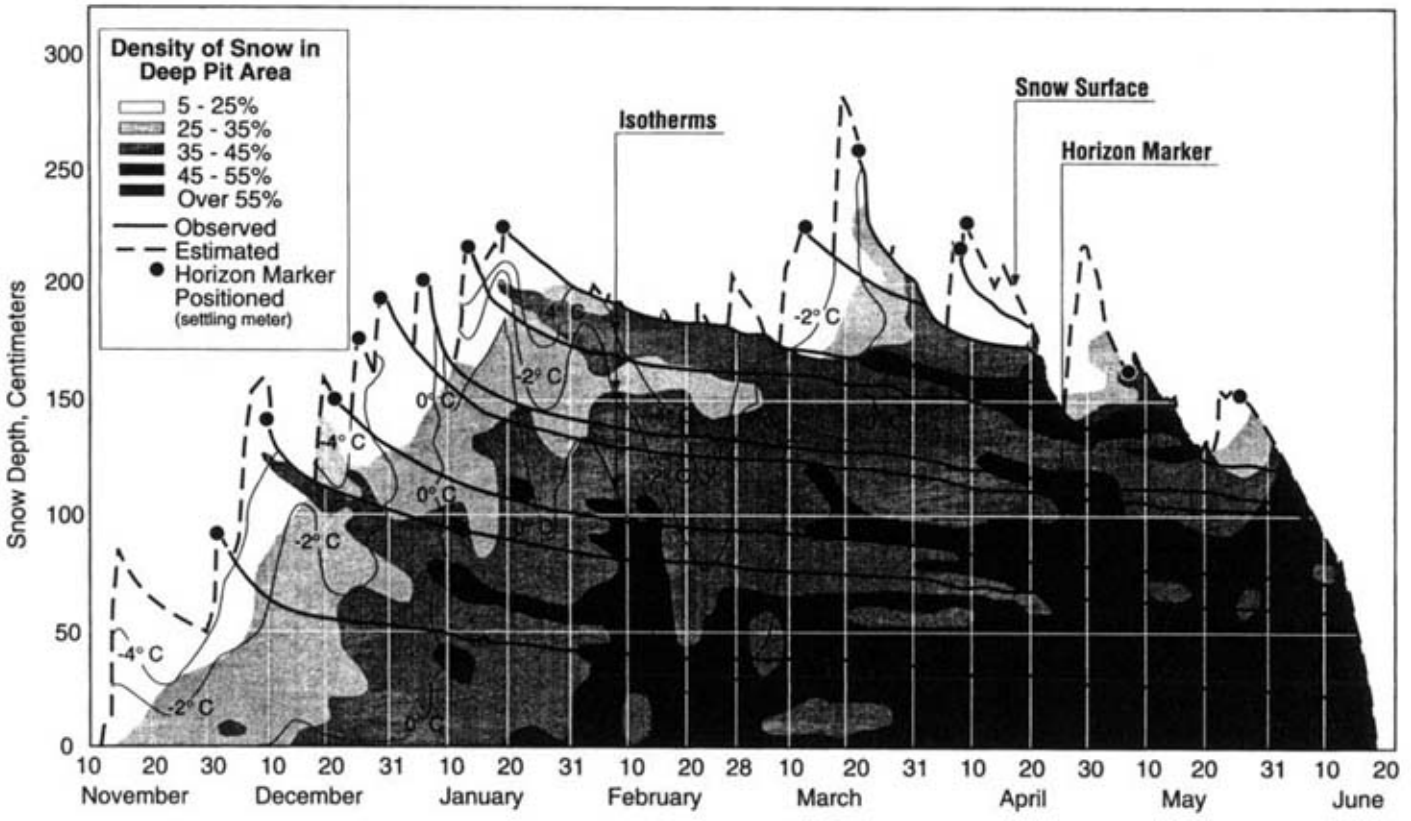


Figure 7.1—Depth, Density, and Temperature Changes in a Snowpack.

established at 80 cal/g (335joules/g), this factor usually must be adjusted to actual snow conditions since the snowpack is not in the form of pure ice at 0°C. The ratio of heat necessary to produce water from snow (and associated free water) to the amount required to melt the same quantity of ice at 0°C is termed the "thermal quality" of the snowpack. For a fully ripe snowpack, the thermal quality can be on the order of .95 to .97.

The rate of snowmelt is dependent upon the many different processes of heat transfer to and from the snowpack, but it is also somewhat dependent upon the snowpack condition. The relative importance of these processes varies seasonally as well as with the day-to-day variation of meteorological factors. The heat transfer processes also vary significantly under various conditions of vegetation, exposure, elevation, and other environmental factors.

The four major natural heat sources in melting snow are: 1) absorbed solar radiation, 2) net long-wave (terrestrial) radiation, 3) convective heat transfer from the air, and 4) latent heat of vaporization by condensation from the air. Two additional minor sources of heat are: 5) conduction of heat from the ground, and 6) heat content of rainwater.

Solar radiation is the prime source of all energy at the earth's surface. The amount of heat transferred to the snowpack by solar radiation varies with latitude, aspect, season, time of day, atmospheric conditions, forest cover, and reflectivity of the snow surface (albedo). The albedo ranges from 40–80%. Long-wave radiation is also an important process of energy exchange to the snowpack. Snow is very nearly a perfect black body with respect to long-wave radiation. Long-wave radiation exchange between the snow surface and the atmosphere is highly variable, depending upon cloud cover conditions, atmospheric water vapor, nighttime cooling, and forest cover. Heat exchange between the snow surface and the atmosphere by convection and by water vapor condensation (items 3 and 4, above) is dependent upon the atmospheric air temperature and vapor pressure gradients, together with the wind gradient in the atmosphere immediately above the snow surface. These processes are particularly important under storm conditions with warm air advection and high relative humidity. In summary, there is no one process of heat exchange with the snowpack that may be universally applied, but the relative importance of each process is dependent on atmospheric, environmental, and geographic conditions for a particular location and a particular time or season. Thorough discussions of the heat exchange processes are given by the USACE (1956), Anderson (1968, 1976), and Male and Gray (1981).

III. DATA REQUIREMENTS, COLLECTION, AND SOURCES

A. Data Requirements

Hydrometeorological data needed for snow hydrology computations include those required for rain-only situations plus the additional variables necessary for the snow accumulation/snowmelt processes involved. These include air temperature data and snow measurements as a minimum but could include wind speed, dewpoint, solar radiation, and other variables if energy budget computations are performed. Since the data requirements for a full energy budget approach are usually unavailable except for hypothetical design conditions and specially instrumented watersheds, analysis and modeling typically use only precipitation and air temperature as input. Table 7.1 summarizes the possible data types that are needed in snow hydrology analysis along with comments on the purpose and application of the data.

Besides the hydrometeorological data requirements stated above, physical data related to the characteristics of the watersheds must be obtained. Area-elevation data, derived from topographic maps or digital elevation models, are of prime importance in mountainous regions. This is particularly necessary for distributed modeling methods that compute hydrometeorological elements on the basis of discretely defined elevation zones or grid cells. Also, data related to the type and density of forest cover, aspect of the watershed elements, exposure to prevailing winds, slopes, etc., are necessary for judging snowmelt parameters for energy budget computations. Overall water balance computations require definition of precipitation variation over the watershed, usually determined from isohyetal maps for the areas involved.

TABLE 7.1. Data Requirements for Snow Analysis.

Data type	Physical process measured or indexed	Analysis application
Streamflow (Q)	a. Continuous discharge b. Runoff volumes	a. Hydrograph analysis, model calibration b. Water supply analysis, forecasting
Precipitation (P)	a. Basin moisture input b. Estimate of SWE	a. Hydrograph analysis, model calibration b. Water supply forecasting
Air Temperature (T_a)	a. Rain/freeze interface b. Index to all energy exchanges c. Factor in energy budget estimates	a. Modeling snow accumulation b. Modeling snowmelt (temp index) c. Modeling snowmelt (energy budget)
Snow Water Equivalent (SWE)	a. Estimate of precipitation b. Index to basin water supply c. Snowpack quantity during ablation	a. Analysis, model calibration b. Water supply forecasting c. Modeling snowmelt
Areal Snow Cover	a. Extent of basin snow cover b. Snow-line elevation	a. Model calibration b. Parameter in forecast models
Snow Fall	a. Estimate of SWE, precipitation b. Accumulation of snow	a. SWE, precip. applications b. Avalanche forecasting
Snow Density	a. Estimate of SWE, precipitation b. Condition of snow	a. SWE, precip. applications b. Avalanche conditions, snow loads
Snow Depth	a. Estimate of SWE, precipitation b. Estimate of weight	a. SWE applications b. Snow load on structures
Snow Albedo	Solar energy absorption	Modeling (energy budget), design floods
Solar Radiation	Solar energy flux	Modeling (energy budget), design floods
Wind Velocity (v)	Factor in estimate of convection/ condensation energy flux and sublimation	Modeling (energy budget), design floods
Dewpoint Temperature (T_d)	Factor in estimate of condensation energy flux and sublimation	Modeling (energy budget), design floods

Air temperature data are quite critical in any modeling or analysis effort, since freezing level must be known during the snow accumulation process to determine the form of precipitation in the basin. Further, air temperature often is used as a key index in computing snowmelt. Historic temperature data are normally archived as maximum and minimum for the day. If a computational interval more frequent than 24 hours is used, then intermediate values may have to be estimated by assuming a typical diurnal temperature pattern (Anderson, 1973). An additional parameter that may be needed in modeling is the lapse rate (vertical temperature profile), which must either be assumed to be a fixed value or estimated from observed temperature readings of atmospheric soundings and ground measurements from stations at different elevations.

Snow data are collected in the form of snow water equivalent on a daily or more frequent basis in the case of automated stations using snow pillows, or monthly (sometimes bi-monthly) in the case of manually measured snow courses. Snow water equivalent data as applied to flood runoff analysis are needed as an independent variable for simplified analyses and seasonal runoff forecasting and as data to assist in calibrating and verifying simulation models. Since snow stations may be the only source of high elevation precipitation data, they also can be used to help estimate basin-wide precipitation input to simulation or statistical models. Data on snow characteristics such as depth and density have only secondary application in snow hydrology but may have importance in other engineering applications such as avalanche prediction and control and snow loading on structures.

B. Data Collection and Utilization

1. Precipitation and Temperature The collection of precipitation data in areas subject to snow accumulation presents additional problems in gaging due to gage freezing, "capping" of the gage by snow, and

high winds. Equipment and field procedures for such conditions are well-documented (Soil Conservation Service, 1972; Male and Gray, 1981). The hydrologist utilizing archived records should, therefore, exercise particular care in assuring that the data are reasonable under these circumstances. When selecting appropriate precipitation, snow, and temperature gages for analysis in a mountainous environment subject to snow conditions, vertical distribution of precipitation and the vertical temperature profile must be considered. The necessary data depends on the application; for simple indexing applications, for instance, a mid-elevation snow gage may be sufficient; however, for detailed simulation, gages located throughout the elevation range are necessary to define the vertical distribution of precipitation and to give a field reference of snow conditions during critical snowmelt times.

2. Snow Water Equivalent The oldest technique for monitoring snowpack is to take manual measurements of snow depth and water content using a snow-sampling tube (Soil Conservation Service, 1972). In the western United States, this type of snow sampling has been done on an organized basis since the 1930's by the Soil Conservation Service; sampling programs also exist in the eastern United States. At a measurement site (or snow course), several measurements are made along a traverse of 90–300 meters in length, each measurement point being 8–30 meters apart. The snow tube is inserted vertically into the snow to the soil surface, the depth is recorded, and the tube and snow core are weighed. The weight of the snow sample is obtained by subtracting the tare weight of the empty tube, and this weight indicates the quantity of water contained in the snow. This is called the "snow water equivalent" (SWE) and is expressed in units of depth. The density of the snow is calculated as the water equivalent divided by the snow depth. The measurements at the sampling points are averaged to obtain single values of depth, SWE, and density for the snow course.

Snow water equivalent is measured at automated sites by a pressure pillow. The pillow is made of stainless steel or a rubber-like material (hypalon) and filled with an antifreeze fluid. It ranges from 2–11 square meters, with the larger sizes required for higher snow accumulation sites. It is connected by a tube to a pressure transducer, which converts the fluid pressure caused by the weight of the snow atop the pillow to an electric signal. The fluid level in the precipitation gage is also sensed by a pressure transducer. The signals from the pillow, precipitation gage, and thermistor (electronic thermometer) are telemetered to a central computer and are available on a daily or more frequent basis. In the western United States, the Soil Conservation Service operates a network of approximately 600 automated snow-monitoring sites in a program called "SNOTEL" (SNOw TELemetry) (Rallison, 1981).

Another method of measuring snow water equivalent is by overflight along flight lines by aircraft. This is done routinely by the National Remote Sensing Hydrology Program of the National Weather Service (Carroll and Allen, 1988). This technique makes use of the fact that there are natural background emissions of gamma radiation from bare ground that are attenuated by water. Snow water equivalent can be calculated from measurements of the gamma radiation from the snowpack compared with that precalibrated from bare ground. Flight lines are typically 16 km long by 300 m wide, and the values reported are mean snow water equivalent over this area. Flight lines exist throughout the northern and western United States as well as in most of Canada.

3. Areal Snow Cover Areal snow cover has been observed visually from the ground over a number of years as an operational guide in assessing snowmelt runoff potential at several locations in the western United States. Beginning in the early 1950's, specially made snow cover observations during the snowmelt season using light aircraft were instituted by the Corps of Engineers for project operations in the Columbia River, and in certain Sierra Nevada drainages in California. These flights continue to be made to provide an assessment of snowcover in conjunction with operational hydrologic forecasts. Remote sensing of snow-covered areas from satellites has also become feasible, while sensing of other snow attributes is still in the research phase. Snow-covered areas for entire basins have been obtained from LANDSAT imagery in various studies and applications. The National Remote Sensing Hydrology Program of the National Weather Service (Carroll and Allen, 1988) reports snow-covered areas both for entire basins and for each of

several elevation zones within selected basins. These data come from the GOES satellite and the NOAA Polar Orbiting Advanced Very High Resolution Radiometer (AVHRR).

4. Other Meteorological Variables Variables such as albedo, dewpoint temperature, solar radiation, and wind velocity are collected at certain first-order weather stations of the National Weather Service and by research laboratories and other institutions. These data are not widely published, so they may have to be obtained by direct contact with the archiving agency. As these data are observed at only a limited number of sites, it is often necessary in an application to use data from sites somewhat removed from the area of interest. In design flood derivations, one is often interested in the general magnitude of these variables and, in particular, the possibilities for extremes. General information of this type can be found in the technical references already cited and in other technical literature on snow hydrology (see Table 7.2).

C. Data Sources

Table 7.2 summarizes data sources typically used in snow hydrology applications. Virtually all data collection made by agencies with official program responsibilities are available in computer databases. Archive files of precipitation, temperature, and other meteorological variables are maintained by the National Weather Service in Asheville, North Carolina, and by NOAA-administered Regional Climate Centers. The largest repository of snow data in the United States is maintained by the Soil Conservation Service in Portland, Oregon. Other agencies that collect snow data and maintain computer data systems include the U.S. Bureau of Reclamation, U.S. Army Corps of Engineers, California Department of Water Resources, and provincial agencies in Canada. In addition to storing the data, the databases allow for data retrieval, report generation, statistical analysis, model calibration, and input data preparation. The capabilities of the databases vary depending on the requirements of the owner agencies. Access permission also varies, as some databases are more restrictive than others.

The various sources for the entire range of hydrometeorological data needed for energy budget calculations are also given in Table 7.2. These include archive data by the NWS and others, as well as research-related data available in books, technical papers, and research laboratories. Four government research laboratories are listed as possible sources of information. The complete hydrometeorological logs from the Cooperative Snow Investigations snow laboratories, active from 1945–1952, are also available. These three field laboratories, located in western Montana, western Oregon, and central California, provided basic data for the report, Snow Hydrology (USACE, 1956).

IV. SNOW ACCUMULATION AND DISTRIBUTION

A. Overview

From the standpoint of general hydrology, the topic of snow accumulation and distribution is a large one, involving considerations such as the physics of snow formation, the effect of terrain and vegetative cover on precipitation and snow accumulation, the effect of interception, and the determination of areal distribution of snow from point measurements; however, in practical engineering applications, the details of the snow accumulation/distribution process are often relatively unimportant compared with the snowmelt runoff determination involved. In a rain-on-snow design flood derivation, for instance, the magnitude of rain may overshadow the snowmelt quantities involved thereby permitting a relatively simple estimate of snow accumulation prior to the design rainstorm. A water supply analysis or streamflow forecast might employ point measurements of snow or accumulated precipitation as indexes to basin-wide SWE with weighting factors determined by model calibration. Although snow accumulation and distribution is not directly used in these examples, the general principles involved are nevertheless invaluable to the practitioner. The details of snow accumulation do come into play when continuous

TABLE 7.2. Data Sources.

Data type	Sources	Contact
Streamflow, Reservoir Storage	a. USGS Water Supply Papers	a. Local USGS office
	b. USGS WATSTOR System	b. Local USGS office for information
	c. Agency Publications, Databases	c. Local/regional agency offices (see text)
	d. CDROM (private vendors)	d. Earth Info, Inc. Hydrosphere, Inc.
Precipitation, Air Temperature	a. NWS Climatological Bulletins	a. Local NWS offices
	b. NOAA National Archives	b. NOAA National Climatic Data Center Federal Building Ashville, NC 28801
	c. State Climatologists	c. State offices (not avail. in every state)
	d. NOAA Regional Climate Centers	d. Contact national center for information
	e. CDROM (private vendors)	e. See above
Snow Water Equivalent	a. SCS Snow Summary Bulletins	a. State SCS offices
	b. SCS SNOTEL database	b. SCS West National Technical Ctr. 101 SW Main St. Portland, OR 97204
	c. State databases	c. State water resources agencies, e.g., California Department of Water Resources 1416 Ninth Street Sacramento, CA 95814
Snow Depth, Snow Fall	a. NOAA Climatological Bulletins	a. See above
	b. NOAA National Archives	b. See above
	c. NOAA Regional Climate Centers	c. See above
Areal Snow Cover	a. NWS Remote Sensing Center	a. Nat'l Operational Hydrologic Remote Sensing Ctr. 6301 34th Avenue South Minneapolis, MN 55450
	b. LANDSAT Photography	b. National Snow and Ice Data Ctr. Cires Campus Box 449 University of Colorado Boulder, CO 80309
Solar Radiation, Dewpoint, Wind Velocity, Albedo	a. NOAA national/regional archives	a. See above
	b. Operational Snow Laboratories	b. (1) USDA Forest Service Central Sierra Snow Lab Soda Springs, CA (2) US Army Corps of Engineers Cold Regions Research & Engineering Lab Hanover, NH (3) USDA Forest Service Rocky Mt. Forest and Range Experiment Sta. Laramie, WY (4) USDA Agricultural Research Service Northwest Watershed Research Center Boise, ID
	c. Textbooks, Handbooks	c. (1) US Army Corps of Engineers (1956) (2) Gray and Male (1981) (3) List (1968) (see also, References)
	d. (1) Proceedings of the Western Snow Conf.	d. (1) Western Snow Conference PO Box 2646 Portland, OR 97208-2646
	(2) Proceedings of the Eastern Snow Conf.	(2) Eastern Snow Conference 161 London Road Newmarket, Ontario L3Y 7A7 Canada
	e. Hydrometeorological logs of Cooperative Snow Investigation Laboratories, 1945-1952	e. US Army Corps of Engineers, North Pacific Div. PO Box 2870 Portland, OR 97208-2870

simulation is used for forecasting or analysis. Here, the vertical and spatial distribution of precipitation and snow must be explicitly modeled during periods of snow accumulation and such factors as interception and sublimation must be taken into account. For comprehensive general coverage on the subject of snow accumulation and distribution, the reader is referred to USACE (1956), Schemenauer et al. (1981), and McKay and Gray (1981).

Hydrologic engineering and streamflow forecasting applications basically require an estimate of SWE for the basin being studied as input into a runoff derivation, and this estimate must either directly or indirectly consider the process of snow accumulation and distribution. The choice of methodology to employ depends upon data availability, the amount of effort to be expended and the type of application involved. Alternative approaches range from simple estimates of a single basin-wide average to detailed simulation using a distributed formulation of the basin. Table 7.3 summarizes three possible approaches of varying complexity.

B. Snow Water Equivalent Estimate from Historic Data

For design flood derivations and other applications, the basin's initial SWE at the beginning of snowmelt can be estimated from historic records of precipitation and snowpack. As noted above, the complexity of the analysis used for this process would logically depend upon the application and the relative importance of snowmelt in the runoff hydrograph. The analysis involved is basically the conversion of point historic data to mean basin or sub-basin SWE and the determination of a specified probability or degree of maximization required for a design flood.

In the simplest applications, snow and precipitation gages in and near the basin can form the basis for relatively rough, single-value estimates of SWE. Normal annual precipitation maps can be used to extrapolate to high, ungaged elevations in mountainous regions. Establishment of a SWE/probability relationship might be based upon frequency analysis of a historic SWE record at a snow station, or it might be based upon precipitation records. When using precipitation data—either station records or isohyets of specified frequency—the determination must be made as to how much precipitation falls and accumulates as snow, to be available at the beginning of the design event. In cold climatic regions or at high mountainous elevations, this would be a simple assumption for a winter period.

A further detailed analysis of initial SWE might include a more thorough determination of the snow "wedge"—the characteristic distribution of snow throughout a range of elevation in a mountainous basin. This type of analysis would employ weighting factors for snow/precipitation stations based upon the extent to which a station represents the total drainage, and it might lead to a distribution of SWE by elevation. A detailed SWE determination might also include a water balance analysis, at which estimates

TABLE 7.3. Alternatives for Estimating Snow Water Equivalent (SWE).

Approach	Possible application	Comment
Simple estimate of SWE	1. Single event rain-on-snow computation	1. Simple estimate based upon historical records. May be adequate where rain dominates
	2. Forecasting in rain-dominated areas	2. Estimates of SWE obtained from real-time data reports
	3. Forecasting seasonal runoff volume using regression methods	3. Measurement provide an index to basin snow
Detailed estimate of SWE, considering elevation and spatial distribution	Design flood derivation, snow-dominated basin	More detailed analysis of historic records required
Simulation of snow accumulation through the accumulation season	1. Detailed design flood derivation	Requires a distributed, continuous-simulation model
	2. Long-range streamflow forecasting	

of basin SWE are developed by balancing all the components of the hydrologic cycle against observed runoff.

In design flood studies, care must be taken in considering the combined probability of the initial SWE in combination with subsequent rainfall and other factors that produce the design flood. A reasonable combination, based upon analysis of historic records, must be used to ensure that the design event is neither under- nor over-maximized. In probable maximum flood determinations involving rain on snow, for instance, the PMP might result from a maximized initial SWE in conjunction with a spring rainfall of moderate magnitude, or the reverse may be true; however, maximizing both would be unrealistic. This question is best addressed by running sensitivity tests and comparing the alternatives.

C. Watershed Definition in Detailed Simulation

Since temperature, and therefore elevation, plays such an important role in defining the conditions of the basin during a precipitation event, the watershed needs to be defined with independent subunits in a detailed snow accumulation analysis. The most common approach is to divide the basin into zones or bands based on elevation. On each band, precipitation, snow, soil moisture, etc., can be simulated independently, as illustrated in Fig. 7.2. Available hydrologic models such as HEC-1 (USACE, 1990) and SSARR (USACE, 1987) provide for the watershed definition to be established relatively easily, employing simplifying assumptions such as defining zone characteristics for the basin through generalized functions. Other models, such as PRMS (Leavesley et al., 1983), divide the watershed into relatively homogeneous, hydrologic response units based on elevation, slope, aspect, and vegetation. (See the section in this chapter regarding computer programs).

D. Simulation of Snow Accumulation

The most thorough approach to estimating SWE is to utilize a continuous simulation model to account for the entire snow accumulation process in relative detail. This requires a distributed basin definition and a frequent (at least daily) calculation time step during which precipitation, temperature, lapse rate, snow line, and runoff are considered. Fig. 7.3 illustrates this process as performed in the SSARR model with a basin subdivided into elevation bands.

V. SNOWMELT ANALYSIS AND SIMULATION

A. Overview of Applications and Approaches

Numerous alternatives present themselves in determining the best approach for computing snowmelt in hydrologic engineering analysis. Ranging from simplified assumptions on discrete storm events to detailed simulation using energy budget principles and distributed definition of the watershed, the choice is dependent upon the application involved, resources available and data availability. Table 7.4 summarizes the options that are possible and how they tend to relate to given types of applications. A typical situation that might be encountered is that of calculating a hypothetical flood from rainfall, either of specified frequency or from a NWS hydrometeorological report. If the meteorological conditions are such that rainfall dominates, and the storm duration is relatively short, it may be quite satisfactory to use a simple approach to estimate snowmelt, e.g., by establishing an antecedent water content with historical analysis, then using an assumed rate of melt or a temperature index applied with a melt-rate factor. The simulation of snow conditioning would not be required since a "ripe" snowpack prior to the storm could be assumed. On the other hand, the derivation of a design flood, or the forecasting of flood runoff in a basin that is snow-dominated, would likely require a more detailed simulation of snow conditioning and snowmelt, perhaps through the use of theoretical or empirical equations as described below. Examples of three applications with differing degrees of complexity appear later in this chapter.

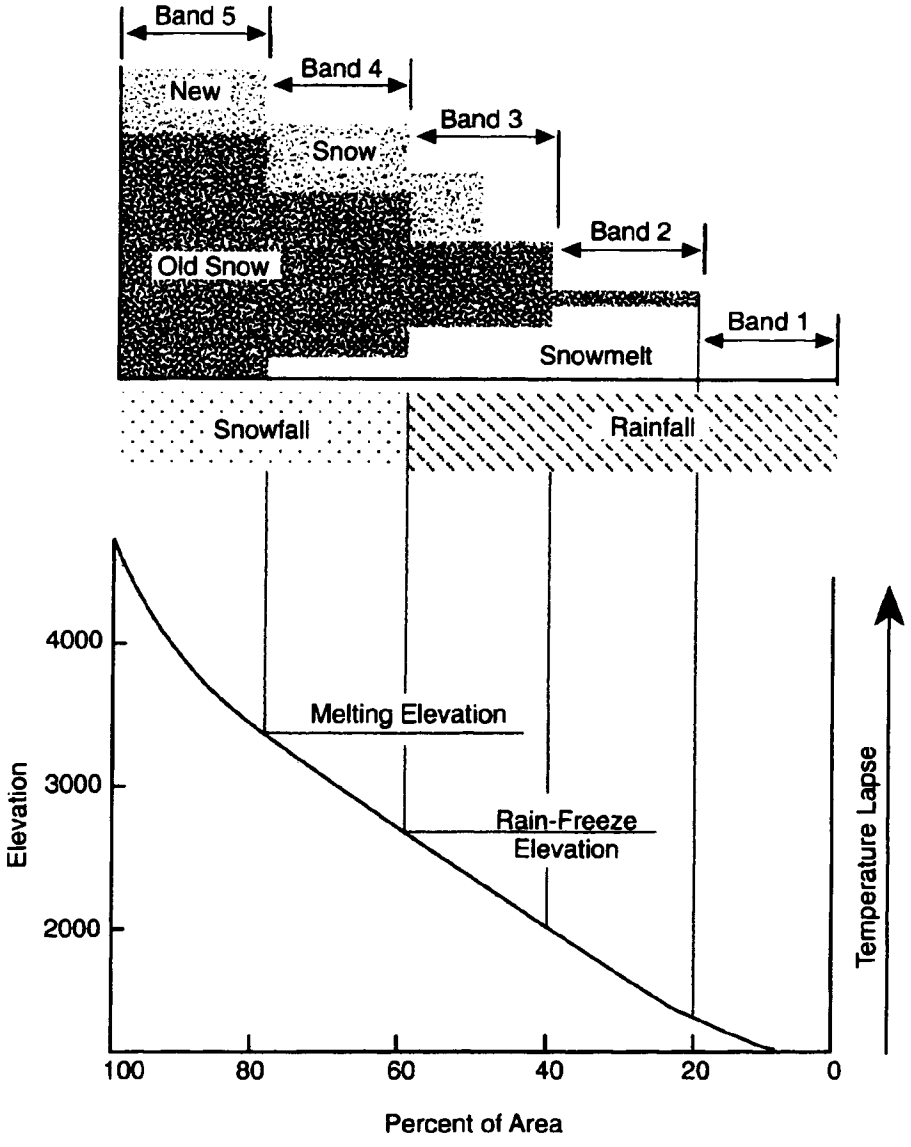


Figure 7.2—Illustration of Distributed Formulation of a Watershed Model using Elevation Bands.

B. Snowmelt

The basic equation defining the amount of energy available for snowmelt can be written as:

$$H_m = H_{sw} + H_{lw} + H_c + H_e + H_r + H_g - d(H_q)/dt \tag{7.1}$$

where H_m is the energy flux available for melt, H_{sw} is the net short-wave radiation flux absorbed by the snow, H_{lw} is the net long-wave radiation flux at snow–air interface, H_c is the convective heat transfer (sensible heat) from the air, H_e is the flux of latent heat released by condensation or sublimation, H_r is the heat flux from rain, H_g is the heat flux from the ground, and H_q is the internal energy content of the snowpack.

While all terms in the above equation are shown as positive, some heat fluxes may in fact be negative, i.e., directed away from the snowpack. For instance, H_{lw} is normally negative in the open and H_e and

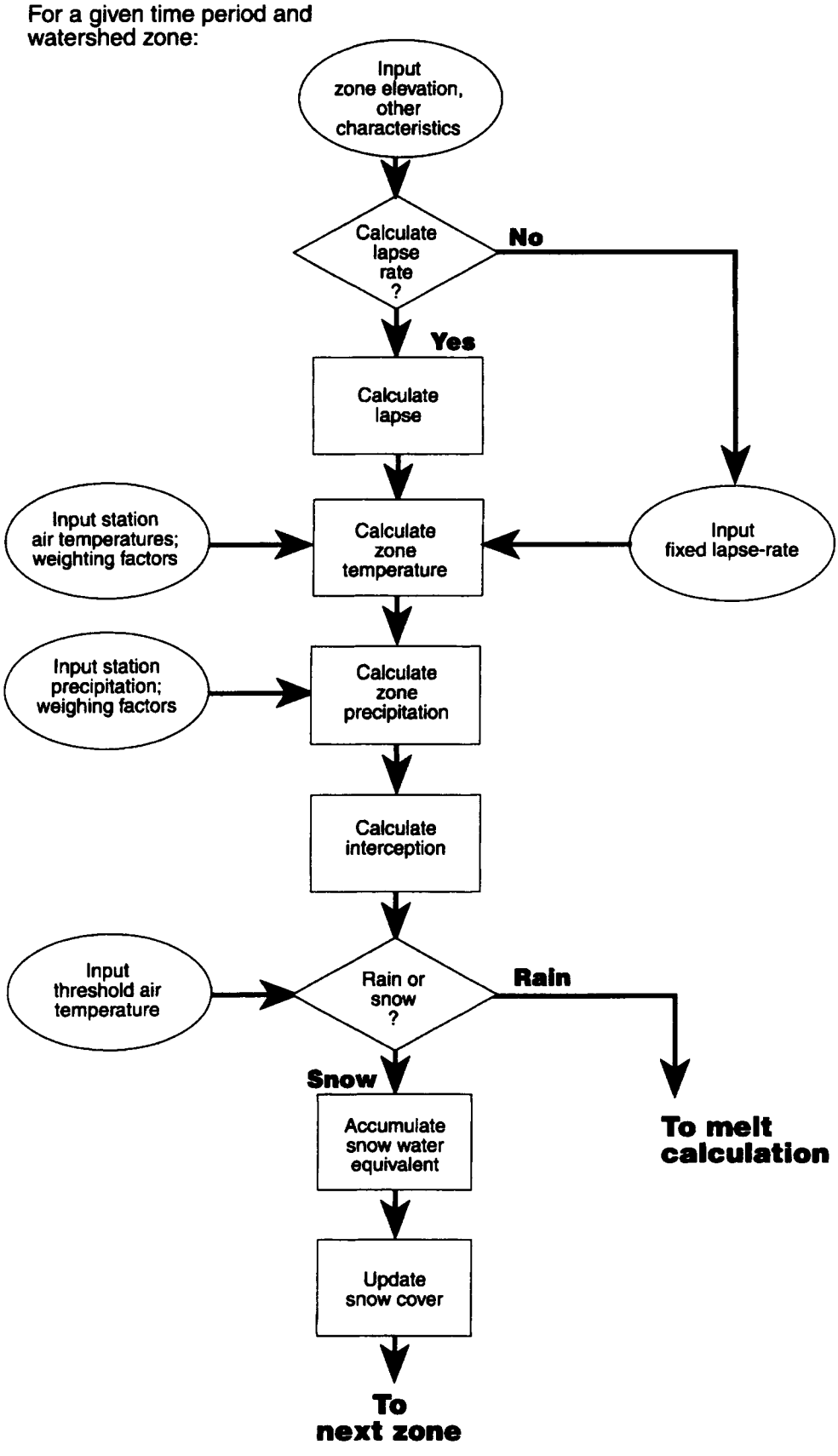


Figure 7.3—Illustration of Snow Accumulation Simulation.

TABLE 7.4. Snowmelt Options.

Application	Example	Basin configuration[1]		Melt calculation			
		Lumped	Distributed	Snow Conditioning	Simplified[2]	Temperature index	Energy Budget
Single-event analysis—Rain-on-snow	Design floods in coastal mountains	Yes	Possibly	Assumed "ripe"	Possibly	Possibly	Possibly
Single-event analysis—Snow (plus rain)	Design floods in interior basins	Yes	Yes	Assumed "ripe"	No	Yes	Yes
Single-event Forecasting Rain-on-snow	Short-term flood forecasting	Yes	Yes	Optional	Possibly[3]	Yes	No
Single-event Forecasting Snow (plus rain)	Short-term flood forecasting	Yes	Yes	Optional	No	Yes	No
Continuous simulation, any environment	Long-term flood & drought forecasting; detailed design analysis	No	Required	Required	No	Yes	Possibly
Detailed simulation in small watersheds	R&D applications; analysis for detailed design; special applications	No	Required	Required	No	No	Yes

[1]Qualitative indicator shown for type of option that might typically be used for application. This is a guideline only. "Yes" or "No" indicates suggested option.
[2]Simplified approach might be to assume a constant or variable moisture input due to snowmelt.
[3]Would be appropriate only in situations where snowmelt is small compared to rain.

dH_q/dt can be either positive or negative. The resulting daily melt produced by a quantity of energy, H_m, is given by:

$$M = H_m/334.9\ B \tag{7.2}$$

where M is the daily snowmelt in mm, H_m is the daily heat flux in kJ/m^2; and B is the thermal quality of the snow. The factor 334.9 is the latent heat of fusion, kJ/kg. As noted previously, the thermal quality of ripe snow is usually between 0.95 and 0.97.

Each of the heat fluxes listed in Equation 7.1 is, in itself, a function of several components, some of which can be difficult to quantify for practical applications. In actual practice then, the theoretical relationships involved are reduced to empirically derived equations that have been found to work satisfactorily in simulation models. Two basic approaches are commonly used: 1) the "energy budget" solution, which employs equations that represent key energy fluxes such as solar radiation, long-wave radiation, wind, heat from condensation of water vapor, etc., and 2) the "temperature index" solution, which uses air temperature as the primary independent variable through the use of a fixed or variable "melt-rate factor." The latter solution is almost exclusively used in practical engineering applications of simulation and analysis for real-time forecasting and reservoir regulation.

1. Energy Budget Equations Numerous researchers have developed equations that represent each of the energy flux components contained in Equation 7.1. These investigations typically include field experiments conducted under a specific physical environment, and often lead to equations that simplify the physical process with empirically derived coefficients. Male and Gray (1981) and Gray and Prowse (1993) contain concise summaries of this work, frequently comparing the findings contained in the literature.

The U.S. Army Corps of Engineers (1956) developed generalized energy budget equations for snowmelt based upon Equations 7.1 and 7.2. These equations employ simplifying assumptions related to the environment in which they are to be used, and make use of variables and coefficients that are feasible in practical engineering applications. The equations were derived primarily for the purpose of developing design floods for major dams, in which prescribed procedures require the use of the combined maximized effect of snow accumulation, snowmelt, storm precipitation, and the effect of the snowpack on runoff. In the case of the maximized snowmelt, there are definite upper limits to the amount of heat exchange by radiative processes or by the advection of heat by the air mass, and these upper limits are determinable only by the use of the energy budget solution. The temperature index solution does not provide the ability to maximize snowmelt rates on a physical basis and, accordingly, it should not be used for design flood determinations for major projects. Two of several generalized equations are presented below with an abbreviated explanation. These and other equations are described in detail in USACE (1956). English unit versions are contained in this reference.

For snowmelt during rain, in which short-wave solar radiation is relatively unimportant and latent heat released by condensation of atmospheric water vapor is relatively high, the following equation applies for open or partly forested areas:

$$M_s = (1.33 + 0.239k_w v + 0.0126P_r)(T_a) + 2.3 \tag{7.3}$$

in which M_s is the total daily snowmelt in mm/day, v is the wind velocity at the 15 m height, in km/hour, k_w is the convection–condensation factor representing the relative exposure of the basin to wind (for unforested areas, $k_w = 1$), T_a is the mean temperature of the saturated air in °C at the 3 m level, and P_r is the daily rainfall in mm.

The constants in the equation are based upon field investigations. The factor 1.33 relates snowmelt due to long-wave radiation to temperature, and the term $0.239k_w v$ represents the effects of convection–condensation melt. The quantity 2.3 accounts for melt from ground heat plus short-wave radiation.

For the case of snowmelt during rain-free periods, direct (short-wave) solar radiation must be accounted for. Several equations were developed by the Corps of Engineers depending upon the degree of forest canopy involved. One, for partly forested areas, is as follows:

$$M_s = k_r(1-F)(2.43I_i)(1-a) + k_w(0.239v)(0.22T'_a + 0.78T'_d) + F(1.33T'_a) \tag{7.4}$$

where M_s is the daily snowmelt in mm/day, T'_a is the difference between the air temperature measured at 3 m and the snow, surface temperature in °C (snow surface temperature is often assumed to be 0°C), T'_d is the difference between the dewpoint temperature measured at 3 m and the snow surface temperature (in °C), I_i is the observed or estimated insolation (solar radiation on horizontal surface) in MJ/m² (one langley = 0.04187 MJ/m²), a is the average snow surface albedo, k_r is the basin short-wave radiation melt factor (it depends upon the average exposure of the open areas to short-wave radiation in comparison with an unshielded horizontal surface, i.e., for a north-facing slope, $k_r < 1.0$, but for a south-facing slope, $k_r > 1.0$), F is an estimated average basin forest canopy cover, effective in shading the area from solar radiation, expressed as a decimal fraction, and k_w is the basin convection–condensation melt factor, as defined above (it depends on the relative exposure of the area to wind). The first term in the above equation is melt due to short-wave radiation. The second term accounts for convection–condensation melt, and the third term represents long-wave radiation melt.

It can be seen that the above energy budget equations have data requirements that may preclude their usage in real-time applications; however, these equations are useful in design flood derivations where variables such as insolation, albedo, etc., can be determined from meteorological tables or upon historic data for the region. These factors would be maximized in the case of a probable maximum flood derivation, for instance. Both of the equations presented above are available in the HEC-1 and SSARR computer programs. The generalized snowmelt equations also provide a useful method of estimating relative magnitudes of melt components. Table 7.5 presents melt quantities calculated from these equations for six hypothetical situations—three with rain, three without.

2. *Temperature Index Solution* Because of the practical difficulties of obtaining data needed for the energy budget equations, common practice is to simulate snowmelt by the "temperature index" solution in real-time forecasting and other applications. This approach utilizes the basic equation:

$$M_s = C_m(T_a - T_b) \tag{7.5}$$

where M_s is the snowmelt in mm per period, C_m is a melt-rate coefficient that is often variable, as discussed below, T_a is the air temperature, and T_b is the base temperature, near 0°C.

Given the variables contained in the energy budget equations previously presented, it can be seen that the employment of temperature only as an index to snowmelt results in further approximation of the actual physical processes involved; yet, considering other uncertainties—particularly in forecasting applications—this does not usually preclude its use from real-time applications.

The melt-rate factor, C_m, is an important key in the successful application of the temperature index equation. While temperature is a reasonably good index of energy flux in heavily forested areas, it is less so in open areas where short-wave radiation or wind velocity play a more important role in the melt process. In clear-weather melt situations, the melt-rate factor varies seasonally, typically increasing as the snowmelt season progresses due to factors such as the decrease in albedo, increased short-wave radiation, longer daylight hours, etc. Because of this, a provision is usually made in simulation models to calculate C_m as a variable, perhaps as a function of accumulated runoff or accumulated degree-days of air temperature.

The choice of base temperature depends upon the computation interval involved and the form of the temperature data involved. If maximum daily temperature is the input variable, then this factor would be higher than 0°C, perhaps as high as 4°C. For a more frequent time interval, the factor would be at or near 0°C.

TABLE 7.5. Relative Magnitude of Snowmelt Factors.

(a) Assumed Conditions

Case	Description	Assumed meteorological conditions				
		T_a	T_d	I_i	P_r	v
1.	Clear, hot, summer day. No forest cover. Albedo = 40%	20	7	30	0	5
2.	Same as Case 1, 50% cloud cover	18	10	20	0	5
3.	Same as Case 1, fresh snow. Albedo = 70%	20	7	30	0	5
—	—	—	—	—	—	—
4.	Heavy wind and rain, warm. No forest cover	10	10	0	80	25
5.	Same as Case 4, but light rain, windy	10	10	0	10	25
6.	Same as Case 5, but light wind	10	10	0	10	5

T_a = Air Temperature, °C
T_d = Dewpoint Temperature, °C
I_i = Solar Insulation, MJ/m²
P_r = Daily rainfall, mm
v = Mean wind velocity, km/hr

(b) Daily Melt Quantities

Case	Snowmelt components, mm					Total Melt, mm	Rain + Melt mm
	M_{sw}	M_{lw}	M_{ce}	M_r	M_g		
1.	43.7	0.0	11.8	0.0	0.5	56.0	56.0
2.	29.2	0.0	14.1	0.0	0.5	43.7	43.7
3.	21.9	0.0	11.8	0.0	0.5	34.2	34.2
—	—	—	—	—	—	—	—
4.	1.8	13.3	47.8	10.1	0.5	73.5	153.5
5.	1.8	13.3	47.8	1.3	0.5	65.7	74.4
6.	1.8	13.3	9.6	1.3	0.5	26.4	36.4

M_{sw} = Short-wave radiation melt
M_{lw} = Long-wave radiation melt
M_{ce} = Convection/Condensation melt
M_r = Rain melt
M_g = Ground heat melt

The possible range of the melt-rate factor can be illustrated by referring to the hypothetical cases presented in Table 7.5. Using the daily melt quantity calculated by the empirical energy budget equations and the temperatures assumed, the melt-rate coefficients calculated with Equation 7.5 would be as shown in Table 7.6. Table 7.6 generally confirms field experience regarding the range in variation of the temperature index melt-rate factor: for clear-melt conditions, the factor varies between 1.5-3.0 mm/°C and increases as the snowmelt season progresses. For rain-melt conditions, the factor can exhibit wide-ranging variations from 2.5–9.0 mm/°C, depending upon wind velocity and, to a lesser extent, the precipitation quantity. These factors would be higher if the temperature index used is the maximum daily temperature. In forecasting practice, the melt-rate factors are estimated through the process of calibrating

TABLE 7.6. Relative Magnitude of Melt-Rate Factors (Refer to Table 7.5 and Equations 7.3, 7.4 and 7.5).

Case	T_a (°C)	T_b (°C)	Melt (mm)	C_m (mm/°C)	Comment
1.	20	0	56.0	2.80	Clear, low albedo
2.	18	0	43.7	2.43	Case 1, cloud cover
3.	20	0	34.2	1.71	Case 1, fresh snow
4.	10	0	73.5	7.35	Heavy rain, windy
5.	10	0	64.7	6.47	Light rain, windy
6.	10	0	26.4	2.64	Light rain, light wind

a hydrologic model. Once established for known historic conditions, the factor can be modified by judgment to be applied to the design condition or forecast situation under consideration. Equations 7.3 and 7.4 can be useful guides in this process. Additional discussion of the magnitude of the temperature index melt-rate factor can be found in USACE (1956), Anderson (1973), and Male and Gray (1981).

C. Snow Condition

As discussed in the overview of physical processes, the storage effect of a snowpack can be a significant factor in determining the time distribution of snowmelt runoff. The condition of the snow—as indicated by its temperature, crystalline structure, density, etc.—will influence the travel rate of meltwater or rain as it moves through the snowpack. In fresh, subfreezing snow, for instance, an initial quantity of meltwater or rain that enters the snowpack will freeze as it warms the snowpack to 0°C. An additional quantity of water is also required to satisfy a liquid water-holding capacity before the snow will release any water by gravity. A further time delay occurs as the free water makes its way to the ground surface, typically by an indirect path around relatively impermeable ice layers. This entire process results in a delay between the onset of atmospheric conditions that cause melt and the corresponding melt runoff. At the same time, it leads to a metamorphosis or ripening of the snowpack. In contrast to a dry, cold snowpack, one that is isothermal at 0°C and ready to yield water to gravity has a relatively minor storage effect, that is associated with the water movement through the snowpack. Male and Gray (1981) comprehensively cover recent research on water movement in a snowpack, including discussion of the propagation rate of the meltwater flux in both homogeneous and stratified snowpacks.

Fig. 7.4 illustrates the storage and transmission of water in the snowpack for an observed rain–on–snow situation. Here, a snowpack having a water equivalent of 680 mm receives input from a 2-day rainstorm. Since the snowpack was colder than 0°C (the beginning snowpack temperature was apparently unknown), some rain/condensation was taken up in raising the temperature of the snow to 0°C through the process of freezing; this amounted to 8 mm of water. An additional quantity of liquid water, 12 mm, was permanently retained in the snowpack, making the total quantity of stored water 20 mm. Finally, the snowpack also temporarily stored some water input as it progressed through the pack. In this case, given the rate of input involved, a time delay of 12 hours occurred between the beginning of rain/melt and the beginning of runoff.

For practical applications in a rain-on-snow situation, snow condition effects can be thought of as an "initial loss" that is subtracted from input, much in the same way as initial losses in dry soil conditions are simulated in rainfall/runoff situations. For the engineer, the problem is to be able to recognize this potential and to be able to incorporate this time lag in the snow hydrology analysis, where appropriate. Practically speaking, this may not be a major factor in design analysis since the snow can usually be assumed to be fully primed prior to the beginning of significant runoff producing melt; however, in certain forecasting situations, the effect of snow conditioning can be noticeable, and it is definitely a factor that must be considered in continuous simulation models that operate through periods antecedent to active snowmelt periods.

1. Cold Content For practical applications, the concept of cold content is useful in quantifying the condition of the snowpack and determining its impact in delaying snowmelt input to the soil. Cold content defines the amount of energy needed to raise a dry snowpack to 0°C, in terms of the amount of water needed to be produced at the surface to release energy by freezing. This can be calculated by:

$$W_c = 0.5\rho_s dT'_s/80 \tag{7.6}$$

where W_c is the cold content (mm), 0.5 is the specific heat of ice (cal/(gm °C)), ρ_s is the average density of the snowpack (g/cc), d is the depth of pack (mm), T'_s is the average temperature deficit of snowpack below 0°C (°C), and 80 is the latent heat of fusion of water (cal/cc).

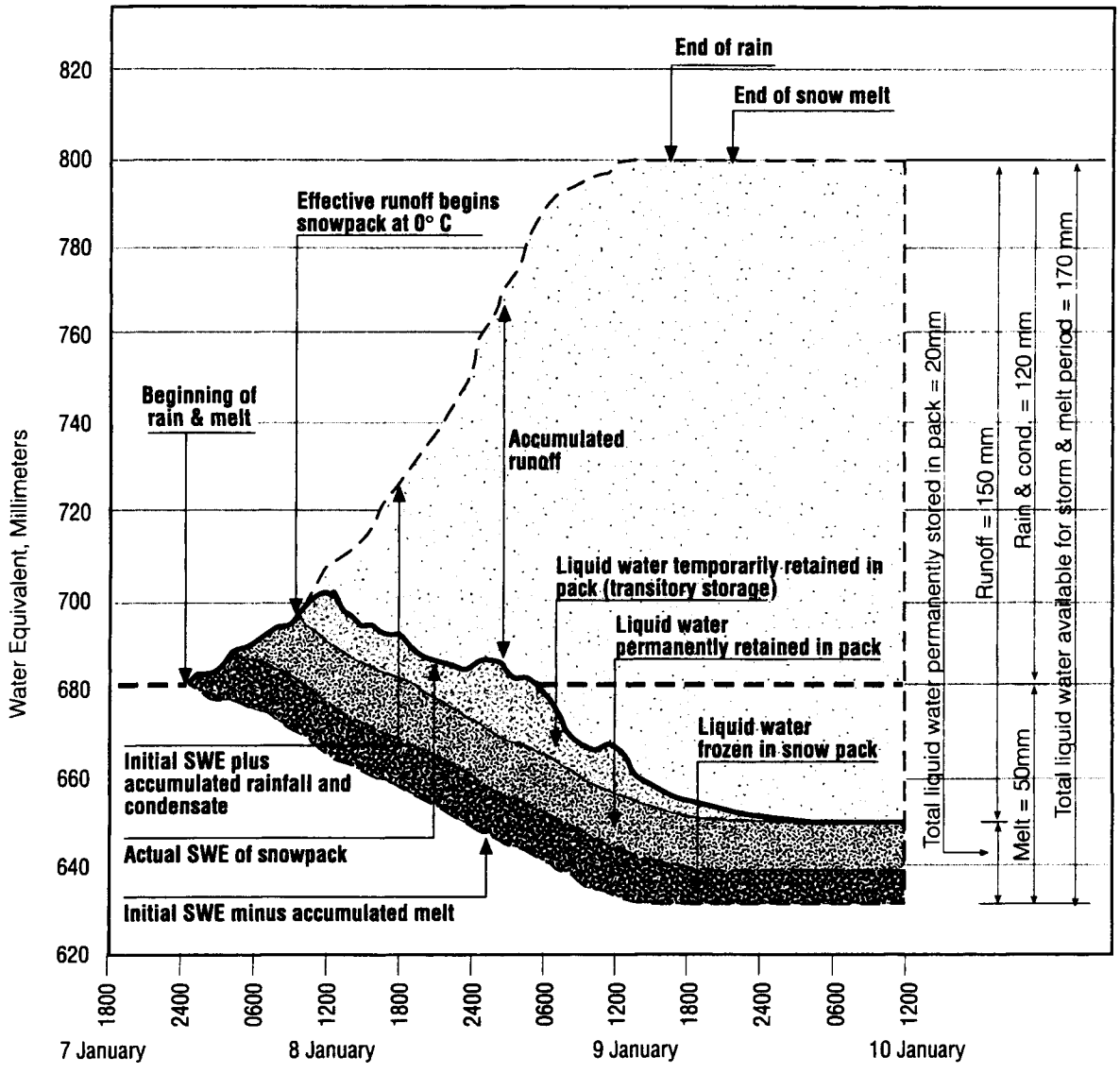


Figure 7.4—Snowpack Water balance during Rain on Snow.

For practical applications, the average snowpack temperature at the time of the application can be estimated based upon the air temperature for 1–3 days prior to the forecast time. Typically, the temperature is close to that of the air at the snowpack surface but approaches 0°C at the ground. For deep snowpacks, a further assumption can be made that only the top 60 cm or so of snow is subject to the influence of air temperature, and that the deeper pack is only 1°C or so below freezing. The density of this layer of snow can also be assumed to be greater than the top layer.

Table 7.7 lists calculated cold content values using Equation 7.6 for various assumed conditions. For the deep snowpack example, the calculation is subdivided into two layers, above and below a 60 cm depth. As can be seen, the cold content is a relatively small factor compared to the potential magnitude of rainfall and snowmelt. It varies from 3% or more of the SWE for "cold" snow to 1% or less of the SWE for deeper snow that is closer to 0°C.

2. ***Liquid Water Holding Capacity*** As shown in Fig. 7.4, the liquid water holding capacity of the snow is a second factor that can be considered an "initial loss" in practical applications of snow hydrology and

TABLE 7.7. Variation in Cold Content.

Descriptive condition	Assumed factors			Calculated factors		
	d (cm)	ρ	T'_s (°C)	SWE (mm)	W_c (mm)	Wc/SWE (%)
Shallow, relatively fresh snowpack. Several days of −8°C temperatures prior to application	40	0.20	6.0	80	3.0	3.8
Same, but warm snowpack	40	0.20	1.0	80	0.5	0.6
Deep snowpack, top 60 cm layer cold	60	0.20	5.0	120	4.0	
	90	0.30	1.0	270	2.0	
	150			390	6.0	1.4
Deep, ripe snowpack. Warmer antecedent conditions	60	0.35	1.0	210	1.0	
	140	0.45	0.5	630	2.0	
	200			840	3.0	0.4

d = snow depth, cm
ρ = snow density
T'_s = average temperature of snow layer, °C below freezing
SWE = beginning snow water equivalent, mm
W_c = cold content from Equation 7.6, mm

forecasting. Unfortunately, very little experimental evidence has led to quantification of this factor; it varies depending upon the depth and density of the snow, the mass of ice layers, and the channelization and honeycombing of the snowpack. It has been observed (USACE, 1956) that at 0°C this factor is approximately 2–5% of the SWE. For most practical applications, a fixed percentage of the SWE is used as an initial loss, in addition to the cold content loss. It should be noted that this magnitude of loss assumes the free drainage of the water. Therefore, in flat areas the snowpack may hold liquid water far in excess of the amount that is associated with mountainous areas.

3. Simulating Snow Condition Since the computation of cold content hinges on the estimation of the snowpack temperature, the common approach to modeling the snow condition in continuous simulation is to maintain an accounting of the relative temperature of the snowpack below freezing as a function of time. In effect, the snowpack is simulated as an energy reservoir; once the reservoir is full (snowpack isothermal and at 0°C), melt water moves to the ground. This can be done through an index relation such as that proposed by Anderson (1973):

$$T_s(2) = T_s(1) + F_p[T_a(2) - T_s(1)] \tag{7.7}$$

where T_s is the index of the snowpack surface temperature at times (1) and (2) (°C), T_a is the air temperature (°C), and F_p is a factor, varying from 0 to 1, representing the relative penetration of the air temperature into the snowpack. If F_p is close to 1.0, the snow temperature remains close to that of the air; thus, values close to 1.0 would be appropriate for shallow snowpacks. For a deep snowpack, a low value of F_p will result in a slow cooling or warming of the snow. The variable T_s is limited to a maximum of 0°C in the simulation process.

Once a snow surface temperature index is established for a computation period, the cold content can be calculated through an equation such as:

$$W_c(2) = W_c(1) + C_r[T_a(2) - T_s(1)] \tag{7.8}$$

where W_c is the cold content at times (1) and (2) (mm), T_a is the air temperature (°C), T_s is the index of the snowpack surface temperature (°C), and C_r is a conversion factor (mm/degree-day). The value of C_r can be made a variable in simulation models by relating it to calendar periods or to a cumulative temperature index function.

D. Snow Accounting During Snowmelt

As snowmelt progresses, the elevation of the snowline increases, and the areal snow cover of the basin decreases. An accounting of this is necessary to be able to: 1) differentiate between snow-free and snow-covered areas, which have different hydrologic characteristics, and 2) determine the elevation of the snowpack for calculating air temperature for indexing melt. A second computation needed is the accounting of the remaining snow water equivalent of the snowpack.

If, in modeling, the basin has been configured into zones of equal elevation as described in the section titled "Snow Water Equivalent Estimate from Historic Data," the accounting of snow cover and quantity can be done on a zone-by-zone basis. One assumption that can be made is to make the zone homogeneous with respect to elevation and either 100% snow-covered or snow-free. This assumption may require a large number of zones for adequate basin representation. Even with a large number of zones, abrupt changes in snowline can occur as a zone changes from snow-covered to snow-free. Because of this, some models provide an ability to simulate a gradual transition within the zone.

An alternative to the distributed approach in accounting for snow during melt is to employ a "snow cover depletion curve" in conjunction with a "lumped" watershed configuration. A snow cover depletion curve describes the basin's snow-covered area as a function of accumulated snow runoff (percent of seasonal total), as shown in Fig. 7.5. Studies have shown this relationship to be of relatively uniform shape for a basin. Using historic field and remote sensing information, a theoretical curve can be developed for a basin. As shown in Fig. 7.5, if actual observations in real time indicate a departure from the theoretical curve, then they follow a proportionally adjusted curve.

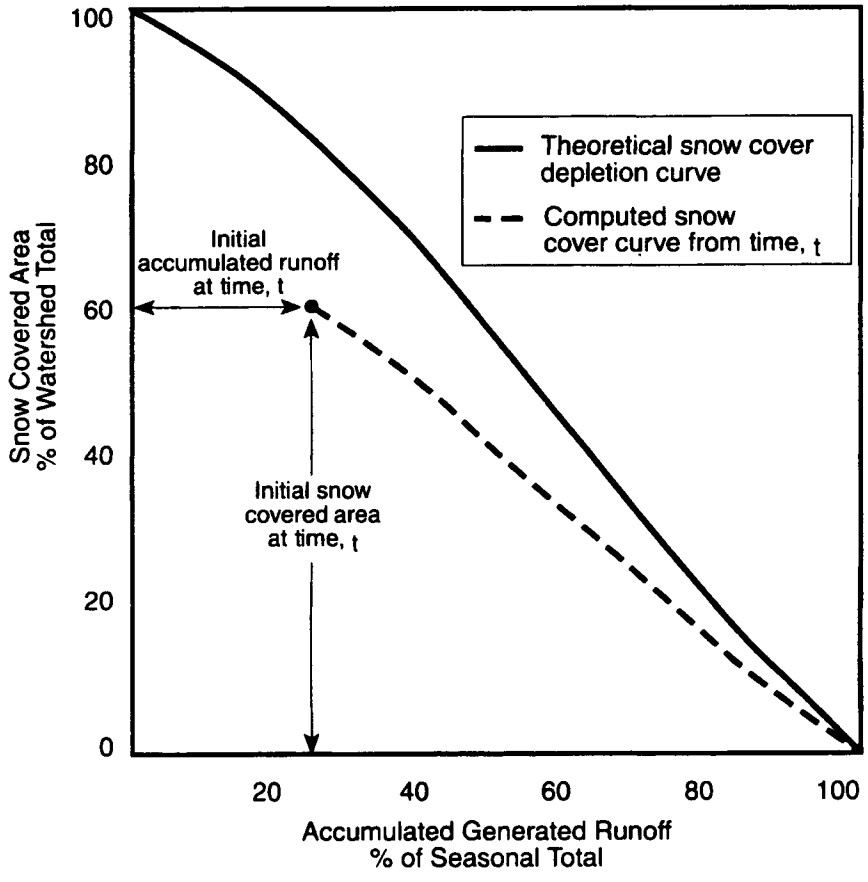


Figure 7.5—Illustration of Snow Cover Depletion Curve.

Chapter 8

FLOODS

I. INTRODUCTION

This chapter covers approaches and analytical processes commonly employed by practicing engineers and hydrologists in flood evaluations. Many publications cover these subjects in great detail. Consequently, this chapter provides an overview of various aspects of floods, a description of flood problems, and a discussion of flood analyses applied to these problems. A greater emphasis is given to the discussion of design flood for water resources facilities including statistical and deterministic methods for flood analysis and probable maximum floods (PMF). Some discussions are devoted to flood hazards and warnings as well as microcomputer software for flood analyses.

A. Flood and Flood Characteristics

A flood may be defined as an event when waters from a river, lake, ocean, or other surface water feature rise above normal limits. Floods often endanger lives and cause millions of dollars in damage. Floods also have beneficial, but seldom recognized, effects such as replenishing soil moisture, cleaning stream channels, depositing fertile fine sediments on agricultural lands, and restricting upstream movement of salt fronts in estuaries.

In general, the definition of floods stated above is insufficient for engineering and related economic, financial, and social analyses. Engineers and hydrologists commonly utilize an "annual flood series" (AFS) in which only the largest event (in terms of peak flow rate or volume) in a year is recognized. One flood is identified in each year even though the value may be smaller than the second or third largest value in some other years. Also, with the AFS on ephemeral streams, it is possible to have a year with a flood magnitude of zero.

Engineers sometimes use a "partial duration series" (PDS) in the analysis. In the PDS, all independent events exceeding a specified base magnitude are identified as floods regardless of the number of events in any given year. The base magnitude for a PDS is commonly established to identify an average of about three events per year as floods. Thus, the PDS provides a greater number of floods for the engineer to analyze.

B. Causes of Floods and Flooding

Flood magnitude and frequency varies spatially and temporally. To evaluate the variability and to suggest appropriate solutions to flood problems requires a recognition of natural and man-made factors

that affect floods. The following paragraphs review major factors that affect river floods and lake and coastal flooding.

1. River Floods

a. Watershed Characteristics. Size, shape, slope, and surface topography are watershed characteristics that affect floods. Larger watersheds produce larger floods of longer durations. These floods generally have lesser year-to-year variation in magnitudes than those from smaller watersheds. In general, a round watershed produces larger flood peaks than a long narrow one. A basin with a steep slope also produces larger flood peaks than basins with small slopes. Larger floods occur on basins with rocky or other impervious surfaces than basins with deep and pervious soil cover.

b. Land Use Characteristics. Forested and heavily vegetated drainage basins usually produce floods of smaller peaks and longer durations than comparable bare basins. The effects of forest fire can significantly increase flood runoff. Construction of reservoirs usually reduces flood peaks downstream; however, small uncontrolled reservoirs on tributaries can disrupt natural runoff timing and cause increase in flood peaks on the receiving stream depending on the positions of these reservoirs. Urban and suburban developments can have profound effects on flooding. Placement of impervious surfaces reduces infiltration. Construction of storm sewers and improvement of drainage channels, including streets with curbs and gutters, causes rapid concentration of runoff. The result is usually a much larger flood peak with increased runoff volume during a shortened flow duration.

c. Precipitation. Rainfall usually is the primary source of flood waters. Intense, short-duration thunderstorm rainfalls often cause severe floods on small streams. Larger streams usually experience the largest floods as a result of longer duration and wider spread storms. In addition to the magnitude, duration, and intensity of rainfall, other factors affect the magnitude and frequency of resulting floods. Seasonally, floods respond differently to comparable rainfalls because rainfall is intercepted by vegetation during growing seasons, infiltration is greater during dry seasons and frozen ground reduces infiltration. While rain falling on snow probably increases total runoff, it may either increase or decrease the peak magnitude.

d. Snowpack. In some parts of the United States, melting of seasonal snowpack provides the primary source of flood waters. Snowmelt rates are controlled by the rates of long- and short-wave radiation, wind, and rain. Since the rates of radiation are limited, peak flood flow rates per unit area are often smaller for snowmelt events than for rain storm events. A snowpack that is saturated and isothermal at the freezing temperature is termed "ripe." Rain falling on a ripe snowpack can add to the runoff and produce a large flood peak; however, rain falling on an unripe snowpack may be absorbed, thereby retarding runoff until the pack ripens (see Chapter 7).

e. Ice. Ice can have a significant effect on floods or flooding. During winter, ice can form in the stream channel to reduce conveyance and cause flooding of overbanks. Spring breakup of ice-covered streams often causes ice jams that temporarily dam the stream resulting in backwater flooding of upstream reaches. Spring flood waters carrying ice are more likely to cause structural damage than ice-free flows.

f. Erosion and Sedimentation. Erosion and sedimentation contribute to a large portion of annual damage losses due to floods. Floods carry sediments eroded from the watershed and scoured from the stream channel. As the flood waters recede, the sediment load is often deposited in the stream channel resulting in the reduction of the flow-carrying capacity of the stream. This can cause flooding of overbanks during subsequent storms.

g. Dam Failure. Though infrequently, dams and other water-retaining structures have failed and produced major floods causing severe loss of lives and properties. Although programs exist for inspection and maintenance of dams, a probability of dam failure remains.

2. *Lake and Coastal Flooding* Climatic and topographic features control the magnitude and frequency of flooding along the shores of lakes, estuaries, and ocean fronts. A principal component of such flooding is "wind setup" or "storm surge" caused by strong, long-duration, on-shore winds and by reduced atmospheric pressure. Wind generated "wave runup" adds to the wind setup or storm surge, often causing severe destruction of facilities along a waterfront. The magnitude of runup is controlled by the on-shore and near-shore beach slopes and the angle of incidence of the waves upon the beach.

Additional factors affecting lake and coastal flooding may include seiche, tides, tsunamis, and sea-level rise. These subjects are discussed in Chapter 10 of this handbook.

C. Measurement of Flood Magnitude

Hydrologists and engineers use several indices to describe flood magnitude. They commonly identify the magnitude of an event by the instantaneous maximum flow rate or flood "peak" in cubic meters per second (m^3/s) or cubic feet per second (cfs). The "peak stage," or maximum water surface elevation associated with a flood peak, is another common index of magnitude. At some stream gages, the instantaneous peak flow rates and stages are not measured. In such situations, the hydrologist or engineer then uses the "maximum daily" flow rate and/or the associated stage. The maximum daily rate is the average rate over 24 hours and is always less than the peak rate. The difference between the peak and the maximum daily value is greater on small streams than on large ones.

Studies comparing flood magnitude between sites of different size in drainage areas commonly use "unit discharge," which is the ratio of peak flow rate to drainage area in m^3/s per square kilometer (km^2) or cfs per square mile (sq. mi.). Studies comparing flood runoff to causative rainfall often use an index of runoff in mm (or inches), which is computed as the ratio of runoff volume for the event to the drainage area. The volume of flood runoff for various durations of time such as 1-, 3-, 7-, 15-, and 30-days is often used in the design of a storage reservoir.

Flood magnitude is sometimes described on the basis of likelihood of exceedance; such as the flood having a 4% chance of being exceeded in any year. The reciprocal of exceedance probability defines an average length of time, in years, between the occurrences of floods of a specified magnitude or larger. Because of this, it has become common to term the flood having a 4% chance of exceedance as the 25-year flood and the flood having a 1% chance of exceedance as the 100-year flood.

Facilities whose failure due to flood could endanger many lives or result in catastrophic economical losses are normally designed for probable maximum flood (PMF), discussed later in this chapter. The likelihood of experiencing a PMF in any future year cannot be properly determined with the current technique in statistical analysis, although it has been assumed to be as small as one in 1,000,000.

D. Flood Hazards

Space occasionally inundated by floods is often ideal for residential and commercial use during nonflood periods. River flood plains provide relatively flat lands where buildings, roads, and utilities can be easily constructed. Alluvial fans, where flood paths are indeterminate, are also popular locations for land developments because they have flatter slopes than surrounding mountain areas. Coastal and lake shores offer recreational and aesthetic opportunities that encourage development. As a result of these characteristics, much development has occurred in areas subject to high flood hazards.

Flood protection works such as storage reservoirs, levees, diversion channels, or channel improvements can protect developments from floods of a selected magnitude; however, complete protection of most development from floods is seldom practical. Flood plain management programs encourage controlled development in flood hazard areas through flood insurance requirements, open space reservations, and building restrictions.

E. Flood Warnings

The National Weather Service (NWS) has been authorized by Congress to provide flood warning services to prevent loss of life and to reduce potential flood damage. Therefore, the NWS has developed flood forecasting models for various watersheds throughout the United States. These models are designed to provide flood warnings based on observed rainfall and runoff information with updates provided as new rainfall and runoff data become available. Although many of the models may be used to estimate the exceedance probability of runoff, the primary concern is to alert communities and individuals of the immediate flood threat by estimating the likely magnitude and timing of flood discharges and/or stages at specific locations. If flood stages are not forecast, they may often be estimated by comparing the forecast discharge with previous flood discharges and their associated stages. Limited resources prevent the NWS from providing flood forecasts for all concerned communities, especially those with small watersheds that have rapid runoff and involve flash flood warnings. Thus, many communities have provided their own flood warning systems, often with assistance from the NWS, other federal agencies and private companies. Flood warning systems are usually combined with emergency evacuation plans, which provide orderly withdrawal from the flood plain prior to significant flooding.

F. Flood Information

Prior to the implementation of the National Flood Insurance Program (NFIP), federal agencies including the U.S. Army Corps of Engineers (USACE), U.S. Geological Survey (USGS), and Tennessee Valley Authority (TVA) provided communities with flood information maps that are similar to the Flood Insurance Rate Maps (FIRMs) of the NFIP. These maps contain information on floods of various exceedance probabilities. In some cases, floods are identified by the dates of occurrence or other terms without assigning an exceedance probability. The need for such flood information maps diminished as the NFIP expanded and the production of such maps gave way to the use of FIRMs. A new type of flood information map evolved from the National Dam Safety Program conducted by the USACE. Because many dams were found to be unsafe under criteria used in the inspections, information regarding downstream flooding by dam failure is viewed as important to evacuation planning and to alerting downstream inhabitants of potential flooding risk. Thus, many federal and state agencies have required the preparation of dam failure inundation maps.

There are two important types of dam failure flood maps. The first involves a nonstorm condition with the reservoir full of water and a sudden or even a timed structural failure of the dam. This condition is probably the most serious as there would be little, if any, time to warn the downstream inhabitants. The second dam failure scenario involves its occurrence during a storm event large enough to overtop and cause dam failure. Thus, large spillway discharges would generally prevail prior to the failure and would alert downstream inhabitants and dam owners of the potential for dam failure. Estimating the exceedance probability of dam failure hydrographs is highly uncertain because of many unknowns associated with dam failure; however, a range of exceedance probabilities for dam failure can be estimated when a storm event is the basis for potential failure.

II. FLOOD ANALYSIS

The objective of flood analysis is to provide information for beneficial social and economic development in flood prone areas or to determine design floods for various components of a water resources development project. This section discusses basic approaches and general considerations for a flood analysis. A detailed discussion of statistical methods for flood analysis is given in the following section, which is followed by a section on estimating floods from rainfall.

A. Basic Approaches

Flood analyses are generally performed to derive peak flow values or complete flood hydrographs at specific locations on streams. The watershed contributing runoff to the location of interest is determined from topographic maps. The derived peak flows and hydrographs are of limited use unless they are associated with some measure of probability of occurrence or physical meaning of the meteorological conditions associated with the floods. Exceedance probability (such as a 1% chance in any given year) combined with potential damages is the desirable approach for selecting a design flood; however, limitations on available data prevent reliability and accuracy in computing the exceedance probability of extremely rare flood events. Because of this, the concept of hydrometeorology was applied as an alternative approach in an attempt to obtain more consistency and reliability in the design of major structures such as dams. The PMF and standard project flood (SPF) discussed later in this chapter are examples.

B. Design Floods

Design floods are the maximum floods against which the various components of a water resource facility or a system of facilities are protected. In some cases a design flood may consist of a peak flow value, whereas other design floods are defined by a complete runoff hydrograph of one or more storm events.

Some projects may require more than one design flood. For a flood control reservoir project, for example, the safety of the dam may be based on the combined reservoir storage and discharge capability to safely pass the PMF. Another design flood may be the 2% chance exceedance flood for selecting the flood control storage in the reservoir. Still other design floods could be used for determining the minimum elevation of relocated highways and railroads or acquisition of land around the reservoir. Design floods for levees may be of the SPF magnitude or a flood of some exceedance probability with smaller discharge.

Selection of design floods depends on the policies and procedures of the entities involved. Seldom are levees designed for floods greater than the SPF. A series of interior floods or drainage may be considered in sizing the drainage facilities, including gravity pipes, pumping plants, pressure pipes, and ponding volumes. Design floods for channel modifications and/or diversions are generally based on a flood of some specified exceedance probability that would not cause flood damage to the facility or overflow damage along the channel banks. Bridges, culverts, and storm drains are customarily designed to pass a discharge of a specified exceedance probability; however, storage is often provided to reduce the required size of the water passage. In this case, flood routing through the storage is necessary.

As a general rule in selecting design floods, the greater the adverse consequences of exceeding the design flood, the greater the design flood should be. This concept is supported by the notion that a higher degree of flood protection provides reduced tangible flood damage, less threat to loss of life, and other intangible advantages. There has been considerable debate regarding the selection of an inflow design flood for the safety of dams. Many engineering conferences and publications have addressed this subject. General consensus regarding large dams located above highly populated flood plains is that no significant threat of dam failure and loss of life due to floods should exist. If the economics of the project prevent achieving this goal, then construction of the project would be an unwise decision. Less agreement has been reached about the inflow design floods for the safety of dams located above flood plains with little or no development. Use of risk analysis (an economic approach) in the selection of the inflow design flood may appear wise at time of construction; however, future downstream development could seriously alter the dam design concept unless land use regulations are enforced for the areas that would be inundated by the dam's failure.

C. Regulatory Floods

Regulatory floods are design floods designated by regulatory agencies for specific purposes. There are three major activities that involve regulatory floods: the National Flood Insurance Program (NFIP)

managed by the Federal Emergency Management Agency (FEMA), the licensing of non Federal hydropower dams managed by the Federal Energy Regulatory Commission (FERC), and the location of power reactor sites, managed by the Nuclear Regulatory Commission. The primary regulatory flood for the NFIP is termed the base flood having 1% exceedance probability in any one year. The flood plain associated with the base flood is identified on the Flood Insurance Rate Map (FIRM), and communities participating in the program must regulate development in the flood plain to reduce future flood damages. Other floods greater and smaller than the flood of 1% exceedance probability are also identified on the FIRM for informational purposes. FERC's regulatory floods are used to assure that construction of a non Federal hydropower dam will not increase the magnitude of natural floods by peaking the flood hydrograph or by dam failure. An individual flood event cannot define all potential flood hazards at nuclear power reactor sites; therefore, all potential sources and magnitudes of floods must be defined to assure a site with essentially no flood risk. This consideration includes artificial floods that could be created by dam failures or other catastrophic events.

III. STATISTICAL ANALYSIS FOR ESTIMATING FLOODS

A. Frequency Analysis

Flood frequency analysis is an important activity in hydrology. The sizing of bridges, culverts, and other facilities, the design capacities of levees, spillways, and other control structures, and the delineation of floodplains depend upon the estimated magnitude of various flood flows that serve as design values. Flood frequency analysis poses a challenge to hydrologists and engineers because available flood records are generally too short to clearly define the frequency of the large flood flows of interest. In such cases, engineers and hydrologists should do their best to use practical knowledge of the processes involved and efficient and robust statistical techniques to develop estimates of flood risk. In the United States, simple flood frequency techniques based primarily on the flood record available at a site are generally restricted to estimating flood flows exceeded with a chance of no less than one in 100.

1. Basic Data and Adjustment for Regulation Streamflow data are collected and stored by many state and federal agencies in the United States. Most cooperate with the U.S. Geological Survey (USGS) in operating a National Water Data Exchange (NAWDEX). A convenient way to access these files is through NAWDEX assistance centers located throughout the United States (contact NAWDEX, U.S. Geological Survey, 421 National Center, Reston, VA, 22092; tel. 703–648–6848). The USGS and its cooperators collect streamflow records at several thousand sites nationwide. Many of these records are published in annual USGS Water Data Reports for each state and are available in major technical libraries and from the National Technical Information Service (NTIS) as well as on CD-ROMs available from private vendors for use with personal computers. Records are stored in the computerized files of the National Water Data Storage and Retrieval System (WATSTORE). The WATSTORE peak-flow records contain annual maximum instantaneous flood-peak discharges and stages, and their dates of occurrence, as well as associated partial duration series. USGS district offices also develop and publish regression relationships for predicting flood quantiles, which are discharges associated with specified exceedance probabilities, for ungaged sites.

If the streamflow record for a site is subject to upstream regulation, the observed flows must be corrected for those activities as well as the effects of watershed changes including urban development, channel modifications, engineering works, and other disturbances.

2. Statistical Terminology The annual peak flood flow is generally thought of as a random variable that can be described by a probability distribution. The parameters of such a distribution are often

estimated with the data available for a site. The flood discharge exceeded with probability q is called the 100q% exceedance probability flood, or 100q% chance flood.

The annual exceedance probability of a flood has often been described by its return period T, or the recurrence interval, where

$$T = 1/q. \tag{8.1}$$

The return period concept can be understood in several ways. In particular, the annual 100q% chance flood is exceeded on average once in a T = 1/q-year period, if the distribution of floods does not change over that period. Likewise, the average time between exceedances of the T-year flood is T years; however, this does not mean that one and only one T-year flood will occur every T years. The probability of the T-year flood being exceeded is 1/T in each and every year. To avoid confusion and improve understanding, the reporting of return periods should be avoided. It is clearer to refer, for example, to the 1% exceedance probability flood, the 1% chance flood, or the flood exceeded with a chance of one in 100, rather than to the 100-year flood.

The general character of a distribution can be described by its mean, denoted by the Greek letter μ, its variance, denoted by σ^2, and its coefficient of skewness, denoted by γ. The mean and variance are defined as

$$\text{mean} = \mu = E[X]$$

$$\text{variance} = \sigma^2 = E[(X-\mu)^2] \tag{8.2}$$

where E[] denotes expectation.

Dimensionless measures of variability and skewness are

$$\text{Coefficient of Variation} = CV = \sigma/\mu, \text{ and}$$

$$\text{Coefficient of Skewness} = \gamma = E[(X-\mu)^3]/\sigma^3. \tag{8.3}$$

Fig. 8.1 illustrates the role of the mean, standard deviation, and skew coefficient in defining the center, scale, and shape of a frequency distribution.

*3. **Probability Distributions*** Many families of distributions have been used to describe flood flows in the United States and in other countries. The normal and lognormal distributions, the Gumbel and Generalized Extreme Value distributions, and the Pearson and log–Pearson type III distributions are discussed below because of their widespread use.

a. Normal and Lognormal Distributions. The normal distribution is useful in hydrology for describing well-behaved phenomena, such as the total annual flow. The bell-shaped normal distribution is symmetric and therefore has a skew coefficient equal to zero. Flood distributions are generally positively skewed so that a normal distribution is not a reasonable model. If Q denotes an annual flood, in many cases the logarithms of flood flows log[Q] are adequately described by a normal distribution. If the logarithms of floods are not quite normally distributed, subtraction of a lower-bound parameter, c, before taking logarithms to obtain log[Q-c] can often resolve this problem, yielding a 3-parameter lognormal distribution (Stedinger, 1980; Hoshi et al., 1984). This distribution is often used to describe floods in Japan and monthly flow series in the United States.

b. Gumbel and Generalized Extreme Value (GEV) Distributions. Annual floods correspond to the maximum flows that occurred within a year. This suggests that their distribution is likely to be a member of a general class of Extreme Value (EV) distributions. If individual daily floods have an "exponential" like

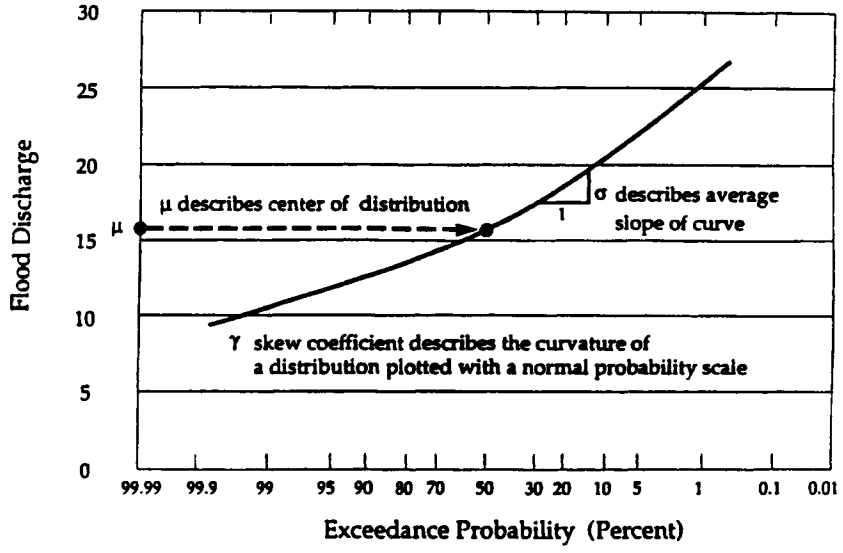


Figure 8.1—The Role of Mean, Standard Deviation, and Coefficient of Skewness in Defining a Flood Frequency Distribution.

upper tail, then the annual maximum is likely to have an Extreme Value type I distribution or Gumbel distribution (Ang and Tang, 1984). Parameter estimation procedures are described by Landwehr et al. (1979) and Phien (1987).

The Gumbel distribution with a fixed skew coefficient of 1.14 is an example of one of many extreme value distributions. The Generalized Extreme Value distribution (GEV) spans three types of extreme value distributions (Gumbel, 1958; Jenkinson, 1969). The GEV distribution is employed in flood studies in the United Kingdom (NERC, 1975; Hosking et al., 1985a,b) and in many other parts of the world (Cunnane, 1988).

c. Pearson Type III and Log–Pearson Type III Distributions. The Pearson type III (P3) distribution is often used to describe hydrologic phenomena. This distribution has a variable skew coefficient, and like the 3-parameter lognormal distribution, has a variable lower bound. For a small coefficient of skewness, it is similar to a normal distribution. When the logarithms of flood flows follow a Pearson type III distribution, the original flows have a log–Pearson type III distribution (LP3). The 2-parameter lognormal distribution is a special case of the LP3 distribution with a log-space skew coefficient of zero. The LP3 distribution is recommended for describing the distribution of floods in the United States in Bulletin 17B by the Interagency Advisory Committee on Water Data (IACWD, 1982) with a regional or weighted estimate of the skewness of the logarithms of the flows. The method for fitting the LP3 distribution is discussed later in this section.

4. Parameter Estimation Fitting a parametric probability distribution to the available flood record provides a compact and smoothed representation of a frequency distribution as well as a systematic procedure for extrapolation of flood risk to discharges beyond the range of the data. Assuming that floods are well described by a particular family of distributions, one needs to estimate the parameters of that distribution so that flood discharges associated with different exceedance probabilities are reasonably approximated by the "fitted" model. The statistical and hydrologic literature contain many methods and philosophies for estimating a distribution's parameters. Perhaps the simplest is the method of moments which computes estimates of the parameters so that the theoretical moments of a distribution match the computed sample moments. The recommended Bulletin 17B procedure for Federal agencies in the United States works with the sample moments of the logarithms of flood flows.

A variation upon the method of moments, which has proved effective in hydrology with the GEV distribution, is the method of probability weighted moments or of L–moments (Hosking et al., 1985a,b).

Estimators of those statistics are linear combinations of the ranked observations and thus do not involve squaring or cubing the observations. As a result they are almost unbiased, whereas the product–moment estimators of the coefficient of variation and coefficient of skewness are both highly-biased and highly-variable (Wallis, 1988; Hosking, 1990).

Another method is the method of maximum likelihood. With this method, parameters are selected that make the fitted distribution most consistent with the available data set in the sense that the selected parameters maximize the probability assigned to the available sample information. Maximum likelihood estimators (MLEs) are known to have good statistical properties in large samples. They are particularly valuable with historical and paleoflood records (discussed later in this section) because of their ability to make efficient use of such data sets (Cohn and Stedinger, 1987). Unfortunately, maximum likelihood estimators are often not available in closed form and thus must be computed using iterative numerical methods. In addition, because MLEs attempt to provide a good fit to all of the observations, they can provide poor estimates of large flood flows when low outliers in a data set are inconsistent with the parametric flood distribution being fit.

Occasionally, modern nonparametric methods are employed to estimate flood-flow frequency relationships (Adamowski, 1985). These methods have the advantage that one does not need to assume that floods are drawn from a particular parametric family of distributions. Modern nonparametric methods have not yet proven themselves in practice and have rarely been used; however, curve fitting procedures that employ plotting positions are simple nonparametric procedures that have a long history of use in hydrology.

B. Selection of a Flood Frequency Distribution

Goodness-of-fit criteria can be used to test whether a selected distribution is inconsistent with a particular data set. Plotting recorded data on the probability paper is an effective way to see what the data looks like and if a fitted frequency curve appears consistent with the available data. Analytical goodness–of–fit tests provides criteria for deciding whether lack of fit is due to sample-to-sample variability or if it reflects a fundamental difference between the fitted distribution and the distribution from which the data were drawn.

Several fundamental issues arise in selecting a distribution for frequency analysis. One could try to determine the true distribution from which the observed flood values were drawn; however, one is unlikely to ever have enough data to answer this question. Moreover, the functional representation of the true distribution of flood discharges would likely have too many parameters to be of practical use. The more practical question is what simple distribution should be used to obtain reasonably accurate estimates of flood flows for exceedance probabilities of interest with the available data sets.

At one time, the distribution that best fit each data set was used for frequency analysis. This approach is not recommended. Employing the best-fitting distribution for each data set provides flow exceedance probability estimates that are too sensitive to sampling variations in the data, and the period of record available. This can result in inconsistencies in flood frequency relationships derived for similar and nearby watersheds. Selection of a family of distributions is best done by a regional analysis using records from many sites. Operational procedures adopted in the United States and by many other countries are based on split-sample/Monte Carlo evaluations of different estimation procedures to find those that yield reliable flood flows and risk estimates for the area in question (Cunnane, 1989, Chapter 6). In the United States, the log–Pearson type III distribution with weighted skew coefficient was adopted based upon a 1974 study by Beard (IACWD, 1982, Appendix 14). In the United Kingdom, the GEV distribution with regional estimators of shape was adopted (NERC, 1975).

1. Plotting Positions and Probability Plots The graphical evaluation of the adequacy of a fitted distribution is often performed by plotting the observations so that they would fall approximately on a straight line if a postulated distribution were the true distribution from which the observations had been drawn. They can also be used to fit a probability distribution to a data set using various curve fitting

procedures. Special commercially-available probability papers can be used for some distributions (normal, 2-parameter lognormal, and Gumbel) that have a fixed shape, perhaps after a logarithmic transformation, or with the more general techniques upon which such special papers are based (Loucks et al., 1981; Cunnane, 1978; Stedinger et al., 1993, p. 18). Special printed probability papers are not available for the Pearson and log-Pearson type III distributions. Lognormal probability paper is commonly used to compare observed flood peaks to a fitted log-Pearson type III frequency curve. When the skew coefficient is not zero, the relationship appears as a curved line (Vogel and McMartin, 1991).

A visual comparison of the empirical and fitted distributions is provided by a plot of the i^{th} largest observed flood $x_{(i)}$ versus the estimated exceedance probability, or probability-plotting position q_i assigned to each ranked flood flow value $x_{(i)}$: $x_{(1)} > x_{(2)} > \ldots > x_{(n)}$ using a probability scale. Fig. 8.2 provides an example of a log-Pearson type III distribution fit to a 24-year record for Fishki Creek at Beacon, New York, using lognormal paper (IACWD, 1982, Figure 12-1). The annual exceedance probability of the i^{th} largest flood in a record containing n values can be estimated by any of the formulas in

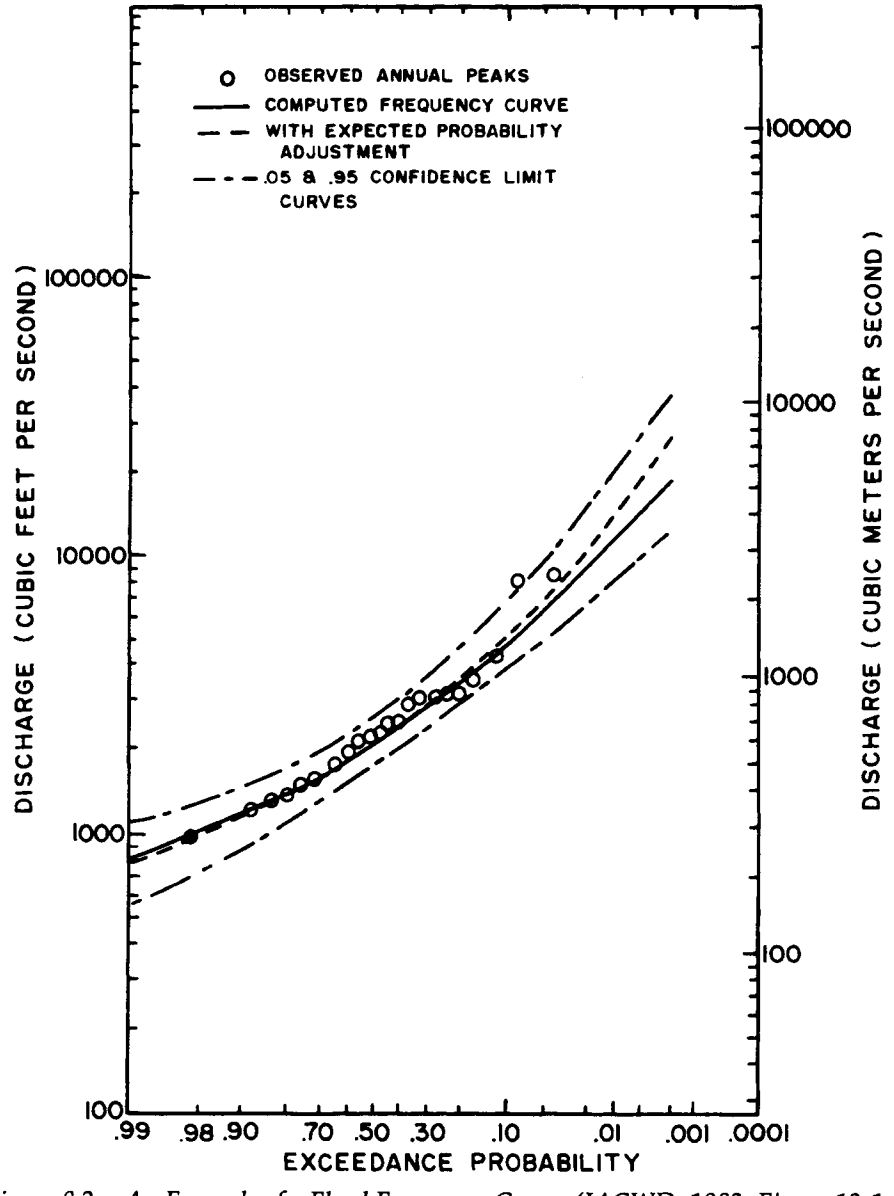


Figure 8.2—An Example of a Flood Frequency Curve. (IACWD, 1982, Figure 12-1.)

TABLE 8.1. Alternative Plotting Positions.

Name	Plotting positions	General comments
Weibull	$i/(n + 1)$	Employed in Bulletin 17B, Unbiased exceedance probabilities
Median	$(i - 0.3175)/(n + 0.365)$	Provides median exceedance probability for all distribution
Blom	$(i - 0.375)/(n + 0.25)$	Unbiased normal quantiles
Cunnane	$(i - 0.40)/(n + 0.20)$	Approximately quantile unbiased
Hazen	$(i - 0.5)/n$	A traditional choice

Table 8.1. Hirsch and Stedinger (1987) review plotting positions for combinations of traditional gaged records and historical information.

Different philosophies lead to selection of different plotting position formulas (Cunnane, 1978). The differences between the Hazen, Cunnane, and Weibull formulas in Table 8.1 are modest for floods that are neither among the largest nor among the smallest; however, they can be appreciable for the largest flood, to which the Weibull formula assigns a chance of one in $(n + 1)$, and the Hazen formula assigns a chance of one in $2n$. Bulletin 17B (IACWD, 1982, p. 26) makes no explicit recommendation. Weibull plotting positions have been used by many federal agencies, as in Fig. 8.2; however, the Cunnane plotting positions are gaining in popularity as a reasonable intermediate choice. The largest three or four events can be plotted with both Hazen and Weibull plotting positions to illustrate the impact of the choice of plotting position formula.

The analyst should remember that the actual exceedance probability associated with the largest observation in a random sample has a mean of $1/(n + 1)$ and a standard deviation of nearly $1/(n + 1)$ (Loucks et al., 1981, p. 109); thus, plotting positions give only crude estimates of the relative range of exceedance probabilities that could be associated with the largest events (Hirsch and Stedinger, 1987). One can use an interval within which the actual exceedance probability falls 90% of the time to represent this uncertainty (Stedinger et al., 1993).

2. *Analytical Goodness-of-Fit Tests* Rigorous statistical tests are available to test whether a set of observations is inconsistent with a particular family of distributions. The Probability Plot Correlation test is often a powerful test of whether a sample has been drawn from a normal, lognormal, Gumbel, Pearson type III, or log–Pearson type III distribution (Filliben, 1975; Vogel, 1987; Vogel and Kroll, 1989; Vogel and McMartin, 1991). Recently developed L-moments can also be used to assess if a Gumbel, GEV, normal, and other distributions are consistent with a data set (Hosking, 1990; Chowdhury et al., 1991). These tests can be employed with records from many sites to construct a more definitive test for a region.

3. *Confidence Intervals* Confidence intervals are a common vehicle for describing the possible value of a parameter or a flood with a given exceedance probability and the certainty with which the value in question can be estimated. Kite (1988) provides equations for calculating such confidence intervals for a large number of distributions. Bulletin 17B (IACWD, 1982, Appendix 9) discusses special procedures for log–Pearson type III distributions; Fig. 8.2 provides an example.

Stedinger (1983b) developed approximate confidence intervals for Pearson and log–Pearson type III quantiles given a known skew coefficient which appears to perform better than the approximate confidence intervals recommended in Bulletin 17B. However, both procedures ignore the uncertainty in the estimated weighted skew coefficient recommended in Bulletin 17B. Chowdhury and Stedinger (1991) show how this omission can be corrected and how confidence intervals that reflect uncertainty in skew coefficients can be constructed.

C. Bulletin 17B Frequency Analysis Method

Uniform procedures for flood-frequency analyses by United States federal agencies are recommended by the Interagency Advisory Committee on Water Data in Bulletin 17B (IACWD, 1982). These methods

evolved over a long period (Thomas, 1985) and were adopted based upon an examination of alternative flood-frequency procedures available in the mid-1970's. The Bulletin 17B procedure estimates two parameters from the data available for a site and a weighted skew coefficient using at-site and regional information. Subsequent research has demonstrated the advantages of making greater use of regional hydrologic information so as to reduce the number of parameters estimated with short at-site records. In particular, use of recently developed index flood procedures for a gaged site with records of modest length is likely to provide more reliable and consistent estimates of flood quantiles (Wallis and Wood, 1985; Landwehr et al., 1987; Potter, 1987, Potter and Lettenmaier, 1990); however, Bulletin 17B procedures are now firmly established in the United States and are not as dependent on the availability of regional analyses.

Bulletin No. 15, "A Uniform Technique for Determining Flood Flow Frequencies," released in December 1967, recommended the log–Pearson type III distribution for use by United States federal agencies. The original Bulletin 17, released in March 1976, extended Bulletin 15 and recommended the log–Pearson type III distribution with a regional estimator of the skew coefficient for the logarithms of the flows. Bulletin 17A followed in 1977, and Bulletin 17B was issued in September 1981 with corrections in March 1982 (IACWD, 1982).

Bulletin 17B prescribes special procedures that address unusual situations, zero-flows or low outliers, historic peaks, regional information, confidence intervals, and expected probabilities. Some features of Bulletin 17B are described here. Those who wish to employ the full Bulletin 17B procedure should refer to that publication or other sources (such as Chow et al., 1988; or Cudworth, 1989).

*1. **Bulletin 17B Moment Estimators*** Bulletin 17B describes procedures for computing flood-flow frequency curves using annual flood series with at least 10 years of data. The recommended technique is to fit a Pearson type III distribution to the base 10 common logarithms of the peak discharges. The flood flow estimator Q_q associated with exceedance probability q is then

$$Log\,[Q_q] = \overline{X} + K_q\,S \tag{8.4}$$

where $\overline{X}$ and S are the sample mean and standard deviation, respectively, and K_q is a frequency factor that depends upon the skew coefficient and selected exceedance probability. Values of K_q for a specified skew coefficient can be read from tables in Bulletin 17B.

For $0.01 \leq q \leq 0.99$ and $|\gamma| < 2$, a good approximation of K_q is:

$$K_q = \left[\frac{2}{\gamma}\left(1 + \frac{\gamma}{6}\,Z_q - \frac{\gamma^2}{36}\right)^3 - \frac{2}{\gamma}\right] \tag{8.5}$$

where Z_q is the quantile of a standard zero-mean unit-variance normal distribution exceeded with probability q.

The flood quantile estimator depends on estimators of the mean, variance, and coefficient of skewness of the logarithms of the flood flows. These estimators (denoted $\overline{X}$, S^2 and G) are calculated as:

$$\overline{X} = \sum_{i=1}^{n} X_i / n$$

$$S^2 = \left[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\right]$$

$$G = n\,\frac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{(n-1)\,(n-2)\,S^3} \tag{8.6}$$

where X_i is the common logarithm of the i^{th} annual peak discharge and n is the number of years of observation.

2. Generalized Skew Because of the variability of at-site sample skew coefficients in small samples, Bulletin 17B recommends weighting the station skew with a generalized skew, which is a regional estimate of the skew coefficient for sites in a region. A generalized skew G_g for sites in the United States can be read from Plate I in the Bulletin. Bulletin 17B indicates that the mean square error (MSE) of the weighted skew estimator is minimized by weighting the station and generalized skew inversely proportional to their individual mean-square errors (IACWD, 1982).

3. Outliers Bulletin 17B defines outliers to be data points that depart significantly from the trend of the remaining data. Procedures for treating outliers require statistical and hydrological judgment. This is an important issue. Because Bulletin 17B works with the logarithms of the observed flood peaks, a single unusual low value, or several such values, can distort the entire frequency curve, particularly its description of the flood-flow frequency relationship for the large events of interest and of importance. A graphical examination of the data and the fitted curve should be performed to check for such problems.

Bulletin 17B suggests that flood peaks identified as low outliers should be discarded from the record and the conditional probability adjustment discussed below be employed. High outliers are retained unless historical information is found to show that such floods are the largest in an extended period. In that case, the Bulletin recommends a historical weighted moments procedure. More recent research has shown that maximum likelihood procedures can make more effective use of historical information, and that it is advantageous to look for such information even when one does not have high outliers (Cohn and Stedinger, 1987).

4. Conditional Probability Adjustment The Bulletin recommends a conditional probability procedure for frequency analysis at sites whose record of annual peaks is truncated by the omission of peaks below a minimum recording threshold, years with zero flow, or low outliers provided that 75% or more of the record is above the truncation level. Maximum likelihood methods developed for use with historical information are effective when a larger fraction of the observations fall below a detection threshold. A frequency curve can also be fit graphically to the probability plot of the observed annual maximum events above the truncation level using plotting positions based on the total record length.

5. Expected Probability A fundamental and unresolved issue is whether, when estimating the flood exceeded with probability q, an analyst should: a) calculate a level that will on average (given the uncertainty in estimated parameters) be exceeded with probability q, or b) provide an unbiased estimation of the true unknown flood level exceeded with probability q. Unfortunately, with limited length records the two criteria lead to different estimations because of the effect of the uncertainty in the estimated parameters. Bulletin 17B (IACWD, 1982; Appendix 11-4) provides a formula for the probabilities that unbiased estimations of the 100q% chance flood will be exceeded. The Bulletin suggests that for lognormal or log–Pearson distributions those equations can be used to make an expected probability adjustment so that reported flood-flows will on average actually be exceeded with the specified probabilities. Fig. 8.2 provides an example of such an adjustment. Such corrections are commonly employed by the U.S. Army Corps of Engineers in the formulation of flood control projects, but are not generally used by other Federal agencies. Arnell (1989) and Stedinger (1983a) show that while the expected probability correction can eliminate the bias in the expected exceedance probability, the correction generally increases the bias in estimated average annual damages for existing facilities located in the flood plain. Thus, a policy either to make or not to make the expected probability adjustments should reflect the intended use of the derived flood-flow frequency relationship.

D. Record Augmentation

It is often possible to effectively extend a short record using a longer nearby record with which observations in the short record are highly correlated. In particular, the longer nearby series can be used to improve estimates of the mean and variance of the events that occur at the short-record site. For this purpose it is not necessary to actually construct the extended series; one only needs the improved estimates of the mean and variance (IACWD, 1982, Appendix 7). Generally, one needs cross correlations of at least 0.7 to improve the mean, and 0.85 to improve the variance (Stedinger et al., 1993, p. 18.37).

E. Risk from Coincidental Events

A flood near the mouth of small tributaries to larger rivers or reservoirs can result from flooding on either the tributary, the larger river or reservoir, or both (Dyhouse, 1985). The flood profiles above a given elevation on the tributary should represent the combined probability of flooding from both causes. Assuming that the flood events on the tributary and on the larger river or reservoir are independent but not mutually exclusive, the combined risk q_c is the sum of the annual probability of flooding above that elevation due to a flood generated on the tributary's drainage q_f, plus the annual probability of flooding above that elevation from the large river or reservoir q_r minus the product of the two:

$$q_c = q_r + q_r - q_r q_f. \tag{8.7}$$

Independence can be assumed when flooding on the larger river or reservoir usually results from a storm centered outside the tributary drainage area that does not produce significant flooding on the tributary, and flooding on the tributary usually results from storms centered over the tributary drainage area that do not generally produce significant flooding on the larger river or reservoir. This approach produces exceedance probabilities intermediate between 1) that produced by considering only the probability of backwater flooding or tributary flooding, $\max[q_r,q_f]$, depending upon which is greater and 2) that produced by ignoring the possibility of concurrent events so that q_c equals $q_r + q_f$.

1. Risks on Alluvial Fans Similar calculations can be used in the evaluation of flood discharge probabilities on alluvial fans. The Federal Emergency Management Agency (FEMA) has developed procedures for combining the probabilities of flood peaks at the apex of the fan and probabilities for the flow path channel geometry to derive the probabilities for flooding at different locations on a fan (Dawdy, 1979; FEMA, 1991).

F. Analysis of Mixed Populations

A common problem in hydrology is that annual maximum series are composed of events that may arise from distinctly different processes. For example, precipitation may correspond to different storm types in different seasons (such as summer thunderstorms, snowmelt events, winter frontal storms, and remnants of tropical hurricanes). Floods arising from these different types of precipitation events may have distinctly different distributions (Cudworth, 1989). Equation 8.7 can be used to describe the combined risk from two different sources of floods that are described by separate component distributions.

Some important questions are: When is it advisable to model several different component flood series separately? And when is it just as reasonable to model the composite annual maximum series directly? If several series are modeled, then more parameters must be estimated, but more data are available if the annual maximum series (or the partial duration series) for each type of event is available. Modeling the component series separately is most attractive when the annual maximum series is composed of only two or three component series with distinctly different distributions that are individually described by relatively simple 2-parameter Gumbel or 2-parameter lognormal distributions (Waylen and Woo, 1982).

However, in such cases, the maximum of the different components still defines a valid annual maximum series that can be modeled directly.

G. Regional Analysis

Frequency analysis is a challenge in hydrology because sufficient information is seldom available at a site to adequately determine the frequency of rare floods. Moreover, design floods and estimates of flood risk are commonly required at sites for which there is no nearby gaged flood record. In such instances one may employ regional flood frequency relationships derived using gaged flood records from similar locations.

Regional hydrologic information is also useful at gaged sites because of the imprecision of flood-risk estimators based upon records of the length typically available. For example, Bulletin 17B uses a regional map of skewness coefficients to calculate flood risk. Record augmentation discussed earlier is only useful if one has several years of record at the site of interest and an adequate nearby site.

The National Research Council (1988) put forward three principles for hydrologic and meteorological modelings; 1) substitute space for time, 2) introduce more structure into models, and 3) focus on extremes or tails as opposed to or even to the exclusion of central characteristics. One substitutes space for time by using hydrologic information at different locations to compensate for short records at a single site. This is easier to do for rainfall, whose characteristics in regions without appreciable relief should be fairly uniform over large areas. Regional estimates of rainfall distributions can allow the selection of design rainfall depths which can serve as the base for calculation of design floods. Conversion of design rainfall to runoff is particularly attractive for urban basins and watersheds subject to regulation and significant development. In other cases, one can employ regional relationships for flood characteristics derived directly from available flood records with appropriate methods of determining the effects of basin area, geomorphology, and land cover.

Two successful examples of regionalization are the index flood method and regional regression procedures discussed below. Both use available samples for sites in a region to construct relationships useful for estimating flood risk at a particular site. For such a method to be effective, the available sample must be representative of floods at the site of interest after necessary corrections are made. Problems occur if sites included in a regional analysis have watershed characteristics that are significantly different from the site of interest. Even worse, if significant development has taken place or if flood flows have been regulated at the site of interest, then the regional relationships may not be applicable. Sites with such effects are generally not used in deriving the regional relationships.

Where appropriate, flood-regionalization methods have significant advantages over the derivation of design floods from design rainfalls. Flood-regionalization methods are relatively easy to apply and different individuals obtain essentially the same flood flow estimates once the background regional analyses have been performed. There are no questions about the relationships between rainfall and flood exceedance probabilities, nor are assumptions needed about antecedent watershed conditions. The United States has very rich networks of stream gaging stations on which regional flood analyses can be based.

1. Index Flood The index flood procedure is a simple regionalization technique with a long history of use in the United States (Dalrymple, 1960). The basic idea behind the index flood method is that the distribution of floods at every site in a region should be the same except for a scale parameter (the index flood) that reflects the size and runoff characteristics of each basin. At gaged sites, the at-site sample mean can be used to estimate that scale value. At ungaged sites, regional regression methods discussed below are generally used to estimate the mean flood.

The normalized frequency curve, or growth curve, can be estimated in several ways. Regional probability-weighted-moment (PWM) index-flood frequency estimation procedures that employ PWM and L-moments have been studied and often provide more accurate estimates of extreme flood quantiles than

would a single-site analysis (Hosking et al., 1985b). The corresponding index-flood procedure based upon L-moments with the GEV distribution used to describe the growth curve should, in practice with reasonably defined regions, be robust and more accurate than procedures that attempt to estimate two or more parameters with the short-records available at many sites (Potter and Lettenmaier, 1990).

The key to the success of the index-flood approach is the identification of reasonable hydrologically similar sets of basins. One may try to group basins geographically, by altitude or by size, recognizing that hydrologic similarity, not geographic proximity, is the important issue. Regional regression can be employed to predict the normalized quantiles using physiographic characteristics of a basin. Modest variability in the normalized quantiles can be tolerated but can degrade the procedure's results (Lettenmaier et al., 1987).

2. ***Regional Regression for Ungaged Sites*** Regression can be used to derive simple models that predict the values of means, standard deviations, discharges exceeded with various probabilities, and normalized index-floods as a function of physiographic river basin characteristics and other parameters. Such models are often used to predict various flood flows at ungaged sites (Jennings and Thomas, 1992). In a nationwide test, these models did as well or better than more complex procedures (Newton and Herrin, 1982).

Consider the traditional log-linear model:

$$\theta_i = a + b_1 \log(\text{Area}_i) + b_2 \log(\text{Slope}_i) + \ldots + e_i \tag{8.8}$$

where drainage area and slope are two of many possible explanatory variables used to estimate θ. A challenge in analyzing such models is that with available flood records, one only obtains sample estimates, denoted as y_i, of the streamflow statistics θ_i. Thus the error e_i is a combination of: a) time-sampling-error $\text{Var}[y_i]$ in sample estimators of y_i and b) underlying model error (lack of fit). The e_i for different sites are cross-correlated if the y_i are based upon concurrent and cross-correlated flow records. Often these problems have been ignored and standard ordinary least squares (OLS) regression has been employed to obtain reasonable estimators of a model's parameters and its accuracy. Such procedures are satisfactory when the flood records are all sufficiently long.

Stedinger and Tasker (1985; 1986) developed the basic theory for Generalized Least Squares (GLS) regression as a regional hydrologic analysis methodology. GLS effectively assigns different weights to the observed streamflow statistics y_i depending upon the estimated time-sampling error in each statistic $\text{Var}[y_i]$ relative to the accuracy of the regression model. Tasker and Stedinger (1986; 1989) and Potter and Faulkner (1987) illustrate GLS use in regionalization studies for the log-space coefficient of skewness and of flow quantiles using physiographic basin characteristics. Advantages of the GLS procedure include more efficient parameter estimates and an unbiased model-error estimator.

Major advantages of regional regression models are that they are relatively easy to employ; an estimate of flood quantile's precision is provided and they embody explanatory variables that reflect the unique physical characteristics of the watershed of interest.

H. Historical Information and Paleofloods

Available at-site systematic records are the traditional and most obvious source of information on the frequency of floods, but they are of limited length. Another source of at-site information is historical flood records. Historical information includes written and other records of large floods left by human observers: newspaper accounts, letters, and flood markers. Paleoflood information describes the many botanical and geophysical sources of information on large floods which are not limited to the locations of past human observations or man-made recording devices. Botanical data can consist of the systematic interpretation of tipped trees, scars, and abnormal tree rings along a water course providing a history documenting the frequency with which different flood stages were exceeded (Hupp, 1987; 1988).

Recent advances in physical paleoflood reconstruction have focused upon the use of slack-water deposits and scour lines as indicators of paleoflood stages, and the absence of large flows that would have left such evidence (Costa, 1987; Stedinger and Baker, 1987; Baker et al., 1988). Such physical evidence of flood stage along a water course with radiocarbon and other dating techniques can provide a relatively accurate and complete catalog of paleofloods in favorable settings with stable channels. A series of studies has documented the potential value of historical information using maximum likelihood procedures (Jin and Stedinger, 1989; Cohn and Stedinger, 1987).

Bulletin 17B contains a less efficient weighted moment procedure that can be used when all floods that exceed a threshold over an extended historical period have been observed. The procedure assumes that the observed gaged-record floods less than a threshold are representative of the unobserved floods which occurred in the historical period outside of the gaged-record period. The weighted moments procedure tends to de-emphasize a large flood, recorded during a short gaged-record, when such a flood is known to be the largest in an extended historical period.

I. Partial Duration Series

There are two general and equally valid approaches for the statistical modeling of flood series. The annual flood series (AFS) approach employs the largest flood in each year and is the most common. The partial duration series (PDS), or peaks-over-threshold approach, considers all independent flood peaks above a truncation or threshold level (Shane and Lynn, 1964; NERC, 1975; Todorovic, 1978). Flood damage studies that consider the possibility of more than one damage-causing flood occurring in a year (or in a season) need to employ partial duration series models; however, such studies should also consider if adequate time passed between damage-causing floods for repairs to occur or another crop to be planted to avoid double-counting. Care must also be exercised to select only the independent peaks for analysis.

An objection to AFS analyses is that they employ only one flood for each year, regardless of whether the second largest flood in a year exceeds the annual flood of other years. In arid regions, there may be no significant flow during an entire water year and thus no significant annual flood to consider. In humid regions, the number of independent and distinct damage-causing flood events may significantly exceed the number of years of record.

Care should be exercised in the frequency interpretation of partial duration series because an n–year series will generally not contain n values. The general frequency implications of a PDS data set can be investigated by plotting the ordered partial-duration peaks $X_{(i)}$ versus an empirical exceedance frequency $f_i = i/n$ for $i = 1, \ldots, m$, when m peaks were observed in an n–year period. Alternatively, $X_{(i)}$ can be plotted against $\log f_i$ to obtain a more linear relationship. For m greater than n, the exceedance frequency for some floods exceeds unity because that level is exceeded on average more than once per year. The annual exceedance probability q associated with a flood discharge of exceedance frequency f is essentially (Chow, 1964):

$$q = 1 - \exp[-f]. \tag{8.9}$$

In most design applications, analytical methods are used to describe the frequency of events in the partial duration series. Such PDS flood-frequency models generally describe both the arrival rate of floods larger than a threshold level and the probability distribution for those exceedance events. The natural estimator of the exceedance rate of a threshold is simply the number of times it was exceeded divided by the length of record. Using a Poisson model for the arrival of floods and an exponential distribution to describe the magnitudes of those flood peaks that exceed the threshold leads to a Gumbel distribution for the annual flood series (Shane and Lynn, 1964; NERC, 1975). If a more flexible generalized Pareto distribution is used to model flood peaks, then Equation 8.9 yields a GEV distribution for the annual flood series (Davison, 1984; Hosking and Wallis, 1987; Stedinger et al., 1993).

J. Bayesian Risk Analysis

Bayesian inference provides another philosophical approach to risk and frequency analysis. Advantages of the Bayesian approach are that it provides a theoretically consistent way of thinking about statistical decision making, allows the explicit modeling of uncertainty in parameters, and provides a theoretically consistent framework for integrating information from at-site gaged-flow records with regional hydrologic information and data from other sources.

The Bayesian approach suggests that supplemental regional hydrologic information should be incorporated through a "prior probability distribution" to augment the information provided by the gaged record for a site (Vicens et al., 1975; Kuczera, 1982; Stedinger, 1983a). Unfortunately, the actual updating or integrating step is straightforward for only a few simple distributions, and the use of prior distributions is sometimes controversial.

IV. ESTIMATING FLOOD FROM RAINFALL

When flood records are not available or not long enough at the site of interest, engineers often use selected storm rainfalls to compute a flood for the site. This approach is sometimes termed the "deterministic method" when compared to the "statistical methods" discussed in the previous section. Use of rainfall to estimate floods is particularly valuable in the situations described below:

a) If a site is essentially ungaged so that flood records are not available for a flood frequency analysis, a design flood can be estimated based upon available rain gage data or regional rainfall frequency analysis.
b) If proposed or recent channelization, urbanization, or other changes within a watershed significantly affect the relationship between rainfall and runoff so that existing flood records are a poor description of current or future flood potential, then the use of rainfall frequency information with a rainfall–runoff model may be advantageous.
c) When reservoir and other water resource operating plans require the simulation of runoff hydrographs at one or more sites, concurrent hydrographs can be generated by multi-basin rainfall–runoff models using multi-site rainfall data from historical or design storms.
d) When interest is in very extreme flood events for design of proposed dams or for safety evaluation of existing dams, available flood records are generally not long enough to properly estimate the rare flood of desired probability of occurrence. In such situations, the flood is estimated by transforming the design storm estimated generally by a hydrometeorological analysis.

To estimate a flood from rainfall, the temporal and spatial distributions of the storm rainfall over a particular drainage area are analyzed. The resulting rainfall, whether it is recorded or based on a statistical or hydrometeorological analysis, is then transformed into a flood peak or hydrograph. The technique could be as simple as applying an empirical equation, using precipitation as an input to obtain peak discharge as an output, or as complex as employing a watershed hydrologic model to capture the complete physical process of rainfall transforming into runoff throughout the watershed.

As statistical methods for estimating rainfalls of prescribed exceedance probabilities are similar to those for floods discussed in the previous section, only estimation of rainfall based on hydrometeorological analysis will be discussed in this section. The discussion will start with an introduction to precipitation analyses covering synopsis of major storms, spatial and time distribution of rainfall, snow accumulation and snowmelt. It is followed by a brief discussion of antecedent and subsequent storms, baseflow, and transformation of rainfall excesses to floods. Finally, the movement of the flood through the channel and overbank storage in the watershed's drainage system is discussed. Precipitation losses due to infiltration are discussed in Chapters 3 and 6.

A. Synopsis of Major Historical Rainstorms

In a broad sense, all rainstorms result from moisture (water vapor) in the atmosphere being lifted to a higher elevation. Atmospheric processes do the lifting, aided in some situations by topography. The more water vapor that is in the lifted air, the better the chance of greater rainfall. Further, the warmer the lifted air is, the more water vapor it can hold. As the air is lifted, it cools, becomes saturated, and the water vapor condenses into ice crystals or small water droplets. The crystals or droplets then increase in size until the force of gravity overcomes vertical lifting, causing precipitation.

In mountainous regions, it is often convenient to categorize precipitation as 1) due to atmospheric processes alone, or 2) due to terrain effects or orography. The precipitation due only to atmospheric processes is termed "convergence" precipitation. The increase or decrease due to terrain is termed "orographic."

The term convergence is derived from the general process of moist air converging at lower levels and then lifted. Atmospheric pressure patterns over large areas, with high and low pressure centers well placed, are mostly responsible for converging and lifting. Weather fronts, the junction of two different atmospheric air masses, often are responsible for lifting. For example, an approaching cold air mass will "dig" below a warmer air mass in its path, thereby lifting it.

Interaction between air masses of different characteristics (moisture source and quantity, temperature, humidity, etc.) can cause heavy convergence rainfall over large areas. These have been termed "general" storms. On the other hand, thunderstorms, limited in supply of moisture, can produce extreme convergence rainfall for a relatively small area and short time. These are termed "local" storms.

East of the Continental Divide, the most intense small area–short duration rainfalls occur within large area–long duration storm situations. In mountainous western regions, the most intense small area–short duration storms in some cases have occurred quite independent of a large scale atmospheric process or storm system. These "local" storms usually cover less than 1300 km^2 (500 sq. mi.). The storm types "local" or "general" are adopted to some extent for expediency since nature can produce rains which blend or fall anywhere between the two extremes.

Tropical storms or hurricanes, the latter with wind speeds exceeding 65 knots, are noted for producing large area–long duration rainfalls over eastern states from Texas to Maine. These storms have also produced some of the most intense small area–short duration rainfalls for this region. Table 8.2 shows the greatest known United States rain depths for various durations. Many of the storms affecting the values in this table are due to hurricanes.

Tropical storms have large circulations originating over warm oceans around a low central pressure. The lower the central pressure, the greater the winds converge to the center and cause greater vertical lifting and resultant rainfall.

As tropical storms progress inland from the coast, the winds are reduced because of friction over the land surface; however, should such storms have a strong circulation against mountains such as the Appalachian or Catskill Mountains in the United States the rainfall can be increased due to the orographic effect.

Tropical storms are also indirectly responsible for some of the largest record rainfalls at a considerable distance from the coast. For example, the great Hearne, Texas storm of June 27–July 1, 1899 and the Thrall, Texas storm of September 8–10, 1921 (Table 8.2) were associated with moisture supplied by tropical storms. Similarly, tropical storms have provided moisture, if not a triggering mechanism, for heavy rains west of the Appalachians all the way to the Great Lakes.

Another aspect of tropical storms is that the slower the storm moves, the more the rain falls along the storm path. The Yankeetown, Florida storm of September 3–7, 1950 progressed very slowly by looping. This storm is also responsible for some United States maximum observed depths (Table 8.2).

Tropical storms have also been responsible for some of the greatest rains in the southwest states. Such storms moving from near Baja California, Mexico have provided moisture for heavy rainfall into Utah and western Colorado (National Weather Service Hydrometeorological Report - [HMR] No. 49, 1977).

TABLE 8.2. Maximum known observed depth–area–duration data for the United States (average rainfall in millimetres (and inches).

Area	Duration (h)						
	6	12	18	24	36	48	72
26 km²	627a	757b	922e	983e	1062e	1095e	1148e
10 mi²	(24.7)	(29.8)	(36.3)	(38.7)	(41.8)	(43.1)	(45.2)
259 km²	498b	668e	826e	894e	963e	988e	1031e
100 mi²	(19.6)	(26.3)	(32.5)	(35.2)	(37.9)	(38.9)	(40.6)
518 km²	455b	650e	798e	869e	932e	958e	996e
200 mi²	(17.9)	(25.6)	(31.4)	(34.2)	(36.7)	(37.7)	(39.2)
1295 km²	391b	625e	754e	831e	889e	914e	947e
500 mi²	(15.4)	(24.6)	(29.7)	(32.7)	(35.0)	(36.0)	(37.3)
2590 km²	340b	574e	696e	767e	836e	856e	886e
1000 mi²	(13.4)	(22.6)	(27.4)	(30.2)	(32.9)	(33.7)	(34.9)
5180 km²	284b	450e	572e	630e	693e	721e	754e
2000 mi²	(11.2)	(17.7)	(22.5)	(24.8)	(27.3)	(28.4)	(29.7)
12950 km²	206bh	282b	358b	394e	475i	526i	620i
5000 mi²	(8.1)	(11.1)	(14.1)	(15.5)	(18.7)	(20.7)	(24.4)
25900 km²	145h	201j	257k	307k	384i	442i	541i
10000 mi²	(5.7)	(7.9)	(10.1)	(12.1)	(15.1)	(17.4)	(21.3)
51800 km²	102h	152i	201k	244k	295i	351i	447i
20000 mi²	(4.0)	(6.0)	(7.9)	(9.6)	(11.6)	(13.8)	(17.6)
129000 km²	64km	107n	135k	160k	201k	251r	335r
50000 mi²	(2.5)	(4.2)	(5.3)	(6.3)	(7.9)	(9.9)	(13.2)
259000 km²	43m	64om	89k	109k	152p	170p	226q
100000 mi²	(1.7)	(2.5)	(3.5)	(4.3)	(6.0)	(6.7)	(8.9)

Storm	Date	Location of Center	
a	17-18 July 1942	Smethport, PA	
b	8-10 September 1921	Thrall, TX	
e	3-7 September 1950	Yankeetown, FL	Hurricane
i	27 June–1 July 1899	Hearne, TX	
k	13-15 March 1929	Elba, AL	
q	5-10 July 1916	Bonifay, FL	Hurricane
n	15-18 April 1900	Eutaw, AL	
m	22-26 May 1908	Chattanooga, OK	
o	19-22 November 1934	Millry, AL	
h	27 June–4 July 1936	Bebe, TX	
j	12-16 April 1927	Jefferson Parish, LA	
r	19-24 September 1967	Cibolo Ck., TX	Hurricane
p	29 September–3 October 1929	Vernon, FL	Hurricane

Source: World Meteorological Organization, 1986.

Extra-tropical storms associated with well-placed high and low pressure centers and air masses with different characteristics (temperature, stability, sources, etc.) have also produced some of the greatest known rainfalls. A wave moving along a front associated with lower pressure creates circumstances that have produced heavy rainfall.

Orographic rain is most notable along the western slopes of the Sierra Nevada in California and the Cascades in Oregon and Washington. Relatively stable air masses in the cool season regularly strike these mountains and the air is lifted, reducing the temperature and causing condensation and precipitation. Examples of unusually severe storms are the January 19–24, 1943 and the December 18–26, 1955 storms, which produced some of the record rains for large regions of the Sierra Nevada (HMR 37, 1962). In many locations, a reduced rain amount is evident due to upwind barriers. It emphasizes that while the air masses tend to show relatively high stability, considerable convergence rainfall is produced. In fact, thunderstorms have been observed in general storm conditions within regions where vertical lifting by the mountain slopes is great.

Chapter 9

URBAN HYDROLOGY

I. INTRODUCTION

The process of urbanization and its hydrologic effects are widely known. At an individual development site, trees and natural cover (see Fig. 9.1a) are replaced by buildings and impervious cover (see Fig. 9.1b). These physical changes result in a loss of interception and depression storage, a decrease in the potential for infiltration, and a redirection of principal flowpaths. On a larger scale, natural channels are cleared, straightened and, in some cases, lined or replaced by pipes that are more hydraulically efficient. The hydrologic impact of these alterations to the physical watershed can change the timing, magnitude, volume, and duration of a flood and its corresponding runoff hydrograph. Baseflow rates can decrease, flood peaks and volumes can increase, and the timing of runoff can change, including decreases in the time-to-peak. Additionally, because of the larger volumes of runoff, the duration of out-of-bank flows can increase. In terms of the frequency of flooding, urban development can increase the frequency of out-of-bank or local flood magnitudes. There are other consequences of the changes in flood runoff characteristics. Surface erosion of cleared areas can increase substantially, erosion of previously stable channel banks can begin, and loadings of water borne pollutants can increase.

The effect of urbanization on flood peaks depends on the exceedance probability of a storm. For the larger peak discharges, the watershed is usually saturated whether urbanized or not, and flood peak increases tend to be smaller because urban development does not create more runoff, only a more efficient runoff collection system. For smaller storm events, urbanization produces larger peaks primarily because of the increase in the runoff volume. Thus, the relative change in flood magnitude is greater for more frequent events and smaller for less frequent ones. This is illustrated in Fig. 9.1c for a hypothetical flood frequency curve. Fig. 9.1d illustrates the changes in peak flow.

The particular engineering emphasis in urban hydrology can be listed as follows.

a) Urbanization adjustments to flood frequency curves from gaged records and estimation of flood peaks in small ungaged urban watersheds for floodway delineation or flood insurance studies administered by the Federal Emergency Management Administration.
b) Flood peak estimation and design of stormwater conveyance systems, such as storm sewers and highway culverts.
c) Storage routing and hydraulic design of stormwater detention facilities that require a full hydrograph.
d) Comprehensive stormwater management computer models for watershed-wide analysis of subarea impacts on the flood hydrology of an entire basin.

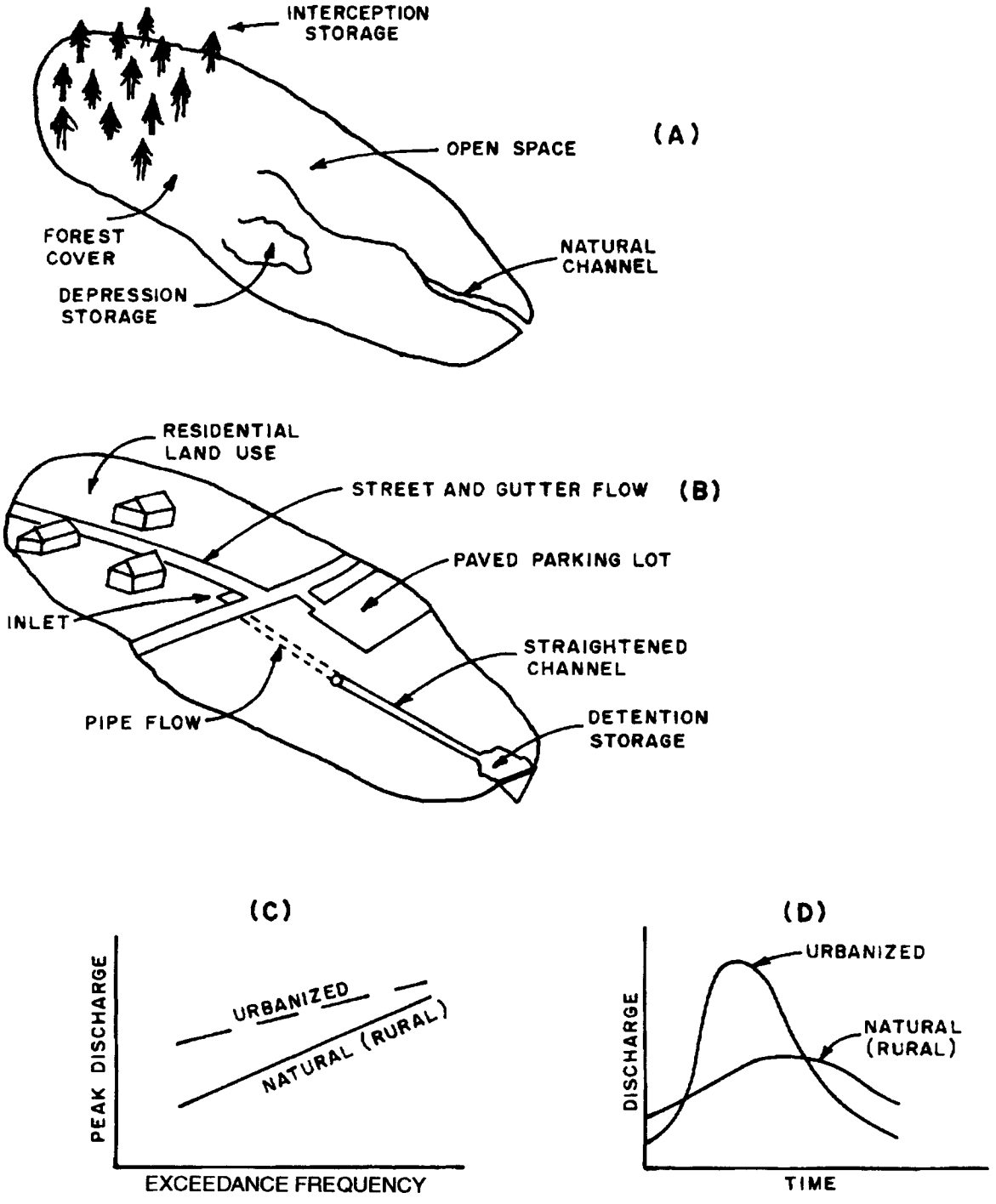


Figure 9.1.—(a) Cover and Storage of a Natural Watershed; (b) Cover and Storage of an Urbanized Watershed; (c) Flood Frequency Curves for Natural and Urbanized Watersheds; (d) Hydrographs for Natural and Urbanized Watersheds.

These engineering and planning applications require that an engineer be knowledgeable regarding rainfall–runoff processes as well as hydraulic principles used in design. Accordingly, this chapter on urban hydrology describes the state-of-the-practice in terms of currently used methods for flood frequency adjustment, precipitation analysis, hydrologic losses, and urban runoff generation. It concludes with two sections on typical urban drainage design calculations and computer model applications. While

the chapter addresses current advances in urban hydrology, it is not intended to be a research document. The reader is encouraged to investigate new research directions through appropriate references which are cited in this chapter. A more detailed overview of each of the major sections follows in the next section.

A. Overview of Urban Hydrology Methods and Processes

A statistical flood frequency analysis is based on the assumption of a homogeneous and time-invariant annual flood record in which rainfall and watershed characteristics remain constant over time. Significant changes in land use are a source of nonhomogeneity in flood statistics. A flood frequency analysis based on a nonhomogeneous record results in inaccurate flood estimates for any return period. Therefore, nonhomogeneities must be identified and the flood record adjusted for the nonhomogeneity prior to making a frequency analysis. An adjustment procedure is provided in this section.

Rainfall information used in urban hydrologic analysis and design most often takes one of two forms: 1) a characteristic rainfall depth and intensity from intensity–duration–frequency curves (IDF) and 2) a time distribution of the design storm depth. These are discussed in the next section on precipitation in the urban watershed.

In addition to changes in flood peak frequency and magnitude, urbanization can cause significant changes to the flood hydrograph. While Fig. 9.1d shows an increase in the peak discharge from the urbanized watershed, other changes are evident, including decreases in the time-to-peak and the base-flow rate, and increases in the runoff volume. These changes to the hydrograph reflect the physical changes to the watershed, with decreases in both interception storage and the potential for infiltration. These changes are reflected in rainfall losses. Methods of accounting for losses on urbanized watersheds are discussed in the section entitled Hydrologic Losses in Developing Watersheds.

The hydrologic methods most frequently used in design work are usually separated into two groups: peak-discharge methods and hydrograph methods. The former is commonly used for design problems on small watersheds where storage effects are unimportant; this would include street drainage, inlet design, and storm sewer design. Hydrograph methods are most often applied to larger watersheds where the effect of storage must be taken into account, such as in the design of regional detention facilities. The fundamentals of these two approaches are discussed in the section on urban runoff methods. Examples of urban drainage design, emphasizing urban storm sewers and detention basins, are presented in the section on urban drainage design calculations. Finally, the application of computer models to problems of storm sewer system design and analysis is contained in a separate section on computer model applications.

B. The Effects of Urbanization on Flood Peaks

A number of hydrologic models are used to evaluate the effect of urbanization on peak discharges. With some models, it is difficult to develop a general statement of the effect of urbanization. For example, with the rational method, urban development would affect both the runoff coefficient and the time of concentration (t_c). The value of t_c is used, in turn, to obtain the design rainfall intensity that appears in the peak discharge formula $Q = ciA$. Thus, because of the many factors involved, it is not possible to make a general statement that a 5% increase in imperviousness will cause an x% increase in the peak discharge for a specific return period.

Other models allow the effect of urbanization to be evaluated directly. For example, a number of regression equations include the percentage of imperviousness as a predictor variable. With such models it is possible to develop a general statement on the effect of urbanization. Sarma et al. (1969) provided one such example:

$$q_p = 2.441A^{0.723}(1 + U)^{1.516} P_E^{1.113} T_R^{-0.403} \quad (9.1)$$

in which A is the drainage area (km^2), U is the imperviousness as a decimal fraction, P_E is the excess rainfall depth (cm), T_R is the rainfall excess duration (hours), and q_P is the peak discharge (m^3/sec). Since the equation has the power model form, the specific effect of urbanization depends on the values of the other predictor variables (i.e., A, P_E, and T_R). However, the relative sensitivity of Equation 9.1 can be used as a measure of the effect of urbanization. The relative sensitivity is given by

$$R_s = \frac{\partial q_p}{\partial U} \cdot \left(\frac{U}{q_p} \right). \tag{9.2}$$

Evaluation of Equation 9.2 yields a relative sensitivity of 1.516 U/(1+U). Thus, at U = 0.1, a 1% change in U (from 10% to 10.1%) causes a change of about 0.1% in the peak discharge. At U = 0.4, a 1% change in U causes a change of about 0.4% in the peak discharge. The importance of U increases as U itself increases. This estimate is an average effect since the actual sensitivity varies with the values of A, P_E, T_R, and U inputs.

Based on the work of Carter (1961) and Anderson (1970), Dunne and Leopold (1978) provided the following equation for estimating the effect of urbanization:

$$f = 1 + 1.5\,U \tag{9.3}$$

in which f is a factor that gives the relative increase in peak discharge for a decimal fraction of imperviousness, U. Thus a 1% increase in U will increase the peak discharge by 1.5%, which is similar to the effect determined from Equations 9.1 and 9.2.

The USGS urban peak discharge equations (Sauer et al., 1981) provide another alternative for assessing the effects of urbanization. The equations are given in Table 9.1. Fig. 9.2 shows the ratio of the urban-to-rural 2-year peak discharge as a function of the percentage of imperviousness and a basin development factor. The peak factors are shown in Fig. 9.3 for the 2-year and 100-year events. For the 2-year event, the ratio ranges from 1 to 4.5, with the latter value indicating complete development. For the 100-year event,

TABLE 9.1. USGS Urban Flood Peak Equations (Sauer et al., 1981).

Regression equations*	R^2	Standard error of regression	
		Log units	Average percent
Seven-parameter equations			
$UQ2 = 2.35A^{0.41}SL^{0.17}(RI2+3)^{2.04}(ST+8)^{-0.65}(13-BDF)^{-0.32}IA^{0.15}RQ2^{0.47}$	0.93	0.1630	±38
$UQ5 = 2.70A^{0.35}SL^{0.16}(RI2+3)^{1.86}(ST+8)^{-0.59}(13-BDF)^{-0.31}IA^{0.11}RQ5^{0.54}$	0.93	0.1584	37
$UQ10 = 2.99A^{0.32}SL^{0.15}(RI2+3)^{1.75}(ST+8)^{-0.57}(13-BDF)^{-0.30}IA^{0.09}RQ10^{0.58}$	0.93	0.1618	38
$UQ25 = 2.78A^{0.31}SL^{0.15}(RI2+3)^{1.76}(ST+8)^{-0.55}(13-BDF)^{-0.29}IA^{0.07}RQ25^{0.60}$	0.93	0.1705	40
$UQ50 = 2.67A^{0.29}SL^{0.15}(RI2+3)^{1.74}(ST+8)^{-0.53}(13-BDF)^{-0.28}IA^{0.06}RQ50^{0.62}$	0.92	0.1774	42
$UQ100 = 2.50A^{0.29}SL^{0.15}(RI2+3)^{1.76}(ST+8)^{-0.52}(13-BDF)^{-0.28}IA^{0.06}RQ100^{0.63}$	0.92	0.1860	44
$UQ500 = 2.27A^{0.29}SL^{0.16}(RI2+3)^{1.86}(ST+8)^{-0.54}(13-BDF)^{-0.27}IA^{0.05}RQ500^{0.63}$	0.90	0.2071	49
Three-parameter equations			
$UQ2 = 13.2A^{0.21}(13-BDF)^{-0.43}RQ2^{0.73}$	0.91	0.1797	±43
$UQ5 = 10.6A^{0.17}(13-BDF)^{-0.39}RQ5^{0.78}$	0.92	0.1705	40
$UQ10 = 9.51A^{0.16}(13-BDF)^{-0.36}RQ10^{0.79}$	0.92	0.1720	41
$UQ25 = 8.86A^{0.15}(13-BDF)^{-0.34}RQ25^{0.80}$	0.92	0.1802	43
$UQ50 = 8.04A^{0.15}(13-BDF)^{-0.32}RQ50^{0.81}$	0.91	0.1865	44
$UQ100 = 7.70A^{0.15}(13-BDF)^{-0.32}RQ100^{0.82}$	0.91	0.1949	46
$UQ500 = 7.47A^{0.16}(13-BDF)^{-0.30}RQ500^{0.82}$	0.89	0.2170	52

*A = basin area (0.2 to 100 sq. mi.); SL = channel slope (3 to 70 ft/mile, with any SL > 70 set to 70); RI2 = basin rainfall depth for 2-yr, 2-hr duration (0.2 to 2.8 inches); ST = basin (detention) storage (0 to 11 percent); BDF = basin development factor (0 to 12); IA = percentage of impervious area (3 to 50%); RQ = rural peak discharge for same exceedance probability (ft^3/sec).

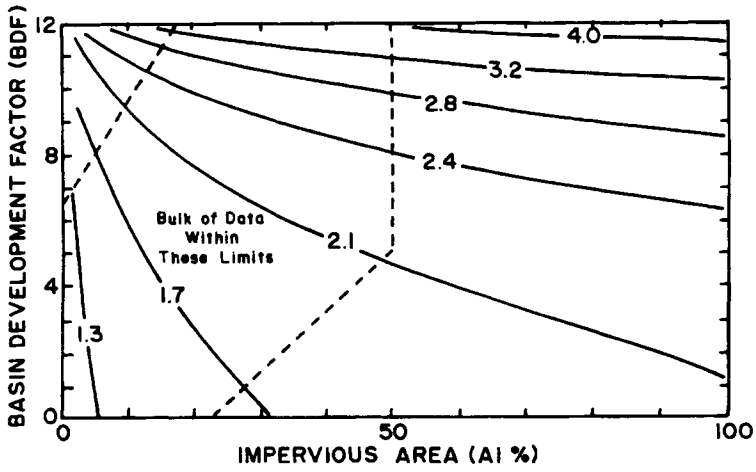


Figure 9.2.—Ratio of the Urban to Rural 2-Year Peak Discharge as a Function of Basin Development Factor (BDF) and Impervious Area (IA) (Sauer et al., 1981).

the ratio has a maximum value of 2.7. For purposes of illustration, assume that basin development occurs in direct proportion to changes in imperviousness. The values of R_s in Table 9.2 will then show the effect of urbanization on peak discharge. The average change in peak discharge due to a 1% change in urbanization is 1.75% and 0.9% for the 2- and 100-year events, respectively. While the methods discussed previously provide an effect of about 1.5%, the USGS equations suggest that the effect is slightly higher for the more frequent storm events and slightly lower for the less frequent storm events. This effect is also evident from the conceptual frequency curve of Fig. 9.1c.

Rantz (1971) provided a method for assessing the effect of urbanization on peak discharges using the simulated data of James (1965) for the San Francisco Bay area. Urbanization is characterized by two variables: 1) the percentage of channels sewered and 2) the percent development of the basin. The percentage of basin developed is approximately twice the percentage of imperviousness. The data in Table 9.2 show the relative sensitivity of the peak discharge to: a) the fraction of imperviousness (U) and b) the combined effect of fraction of channels sewered and basin developed (D_b). For urbanization as measured by the fraction change in imperviousness, the means of the relative sensitivities are 2.6%, 1.7%, and 1.2% for the 2-, 10-, and 100-year events, respectively. When both the fractions of channel sewered

TABLE 9.2. Relative Sensitivity (R_s) of Peak Discharge to Urbanization (U) and Basin Development (D_b) for the USGS Seven-Parameter Equations of Table 9.1 and The Peak Factors of Rantz (1971).

T (yrs)	U	USGS R_s	T (yrs)	U	Rantz R_s	T (yrs)	D_b	Rantz R_s
2	0.20	0.016	2	0.10	0.025	2	0.10	0.040
	0.25	0.016		0.20	0.025		0.20	0.060
	0.30	0.018		0.30	0.026		0.30	0.085
	0.35	0.020		0.40	0.029		0.40	0.100
100	0.20	0.010	10	0.10	0.015	10	0.10	0.022
	0.25	0.008		0.20	0.017		0.20	0.040
	0.30	0.008		0.30	0.018		0.30	0.050
	0.35	0.010		0.40	0.018		0.40	0.092
			100	0.10	0.011	100	0.10	0.010
				0.20	0.012		0.20	0.025
				0.30	0.013		0.30	0.050
				0.40	0.012		0.40	0.055

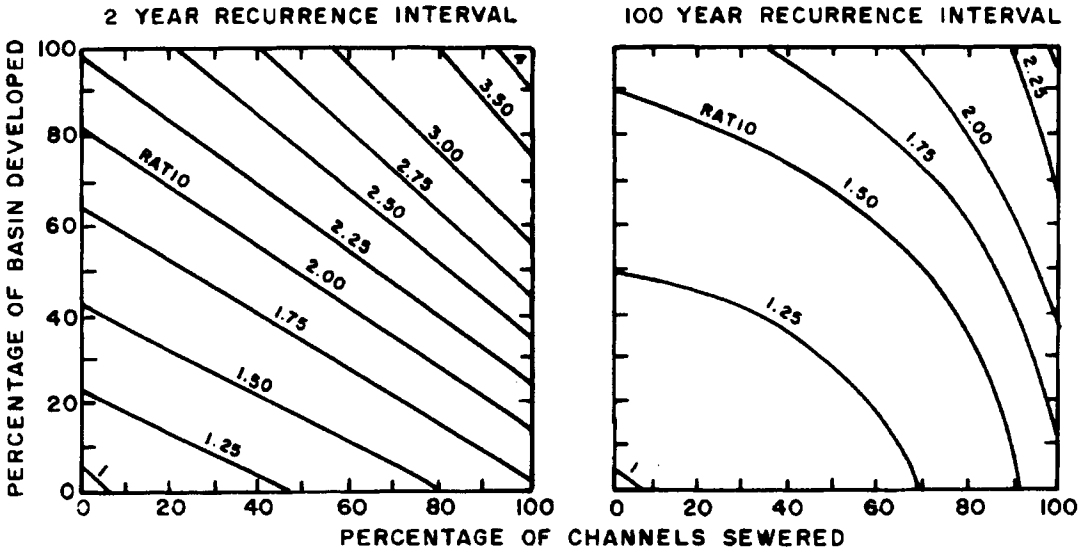


Figure 9.3.—Effect of Percentages of Imperviousness and Channels Sewered on the Ratio of the Urbanized to Undeveloped Peak Discharges (Rantz, 1971).

and basin developed are used as indices of development, the relative sensitivities are considerably higher. The means of the relative sensitivities are 7.1%, 5.1%, and 3.5% for the 2-, 10-, and 100-year events, respectively. These values are larger than the values suggested by the other methods discussed in the preceding paragraphs. The procedures cited above are presented as examples of the techniques available for urban flood frequency analysis. Many other statewide adjustment procedures are available and the interested reader is urged to consult a state office of the U.S. Geological Survey or a State Department of Water or Environmental Resources for the appropriate regional method.

C. A Method for Adjusting a Flood Record

The literature does not identify a single method that is considered the best for adjusting a flood record for the effects of urbanization. Each method depends on the data used to calibrate the prediction model. Unfortunately, the database used to calibrate the models is very sparse. The sensitivities suggest that a 1% increase in urbanization causes an increase in peak discharge of about 1 to 2.5% for the 100-year event and the 2-year events, respectively; however, there is considerable variation at any return period.

Based on the general trends of the available data, the following method is suggested for adjusting a flood record. Fig. 9.4 shows the peak adjustment factor (f_i) as a function of the exceedance probability for percentages of urbanization up to 60%. In Fig. 9.4, urbanization can be defined as the imperviousness or the degree of basin development. The greatest effect is for the more frequent events and the highest percentage of urbanization. Given the exceedance probability or return period of a flood peak for a nonurbanized watershed, one can assess the effect of increasing urbanization by multiplying the discharge by the peak adjustment factor for the corresponding return period and percentage of urbanization. Where it is necessary to adjust a discharge from a partially urbanized watershed to a discharge for another watershed condition, an engineer/hydrologist can divide the discharge by the peak adjustment factor for the existing condition and then multiply the resulting "rural" discharge by the peak adjustment factor for the second watershed condition. The first operation (i.e., division) adjusts the discharge to a magnitude representative of a natural or nonurban condition. The second operation (i.e., multiplication) adjusts the discharge to a newly developed condition. The adjustment method of Fig. 9.4 requires an exceedance probability obtained from a plotting position formula. The following procedure can be used

to adjust a flood record for which the individual flood events have occurred on a watershed that is undergoing a continuous change in the level of urbanization.

1) Identify both the percentage of urbanization for each event in the flood record and the percentage of urbanization for which an adjusted flood record is needed.
2) Compute the rank (m) and exceedance probability for each event in the flood record. The Weibull plotting position formula can be used to compute the probability, where the plotting position (pp) for events given in descending order equals the rank (m) divided by 1 plus the sample size (N); i.e., pp = m/(N + 1) = estimated annual exceedance probability of event having rank m.
3) Using the exceedance probability estimate and the actual percentage of urbanization, one then finds from Fig. 9.4 the peak adjustment factor (f_1) to transform the measured peak from the actual level of urbanization to a nonurbanized condition.
4) Using the exceedance probability and the percentage of urbanization for which a flood series is needed, one finds from Fig. 9.4 the peak adjustment factor (f_2) that is necessary to transform the nonurbanized peak to a discharge for the desired level of urbanization.
5) Compute the adjusted discharge (Q_a) by

$$Q_a = \frac{f_2}{f_1} Q \tag{9.4}$$

in which Q is the measured discharge.

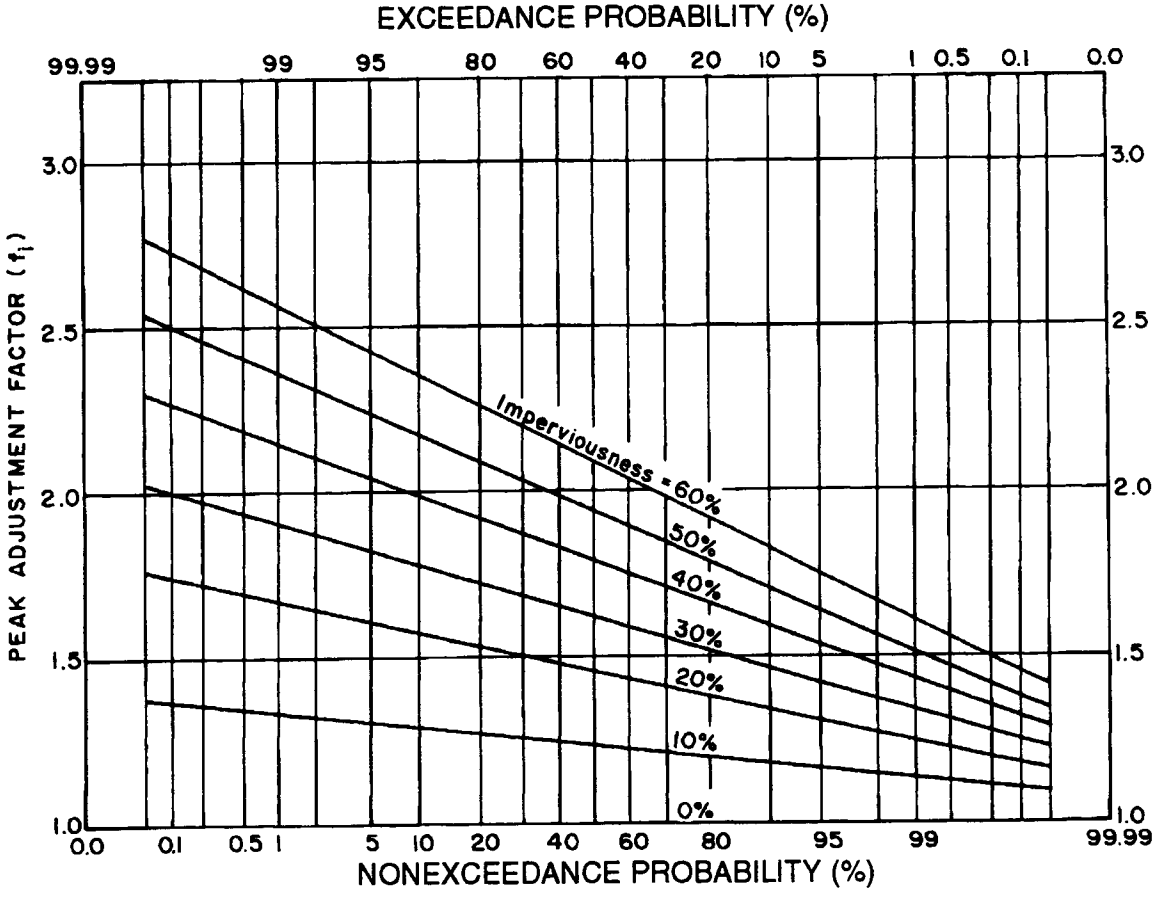


Figure 9.4.—Peak Adjustment Factors for Urbanizing Watersheds (McCuen, 1989).

6) Repeat steps 3, 4, and 5 for each event in the flood record and rank the adjusted series.
7) If there are changes in the ranks of the measured (Q) and adjusted (Q_a) flood series, repeat steps 2 through 6 until the ranks do not change.

The 13-year annual maximum flood record shown in column 3 of Table 9.3 for the Alhambra Wash watershed, Los Angeles, California, is used to illustrate the application of Fig. 9.4. The percentage of imperviousness, used as a measure of urbanization, increased from 26 to 44% over the thirteen-year period (see column 2 of Table 9.3). Ultimately, the percentage increased to 45%. The ranks (column 4) and the Weibull exceedance probability and percentage of imperviousness were used to obtain the peak adjustment factor (f_1 of Equation 9.4) that converts the measured flow to an estimated flow for nonurbanized watershed conditions. The exceedance probability and the ultimate percentage (45%) were used to obtain the peak adjustment watershed condition to the flow for the ultimate development condition. The ratio f_2/f_1 converts the measured flow to a flow for ultimate development. Columns 6 through 8 of Table 9.3 give initial estimates of the adjustment factors and the adjusted discharges. Since the ranks of the adjusted discharges (column 9) are different from the ranks of the measured series (column 4), a second iteration is necessary; this is shown in columns 10 to 14. The final estimates for ultimate conditions are given in column 13. Although the length of the annual series is too short in the example for flood frequency analysis, that would be the next step if the annual series exceeded 20 years of record.

D. Design-Storm Approach

Models such as the USGS regression equations of Table 9.1 are very useful for developing a frequency curve for a watershed. For design problems involving large nonhomogeneous watersheds where watershed subdivision is necessary, the regression-equation approach is inadequate. In such cases, the hydraulic model/design storm approach is the most frequently used method for evaluating the effect of urban/suburban development on peak discharge rates. In this approach, it is necessary for the hydrologic engineer to select: 1) a regional rainfall intensity–duration–frequency (IDF) curve, 2) a design–storm hyetograph, 3) a technique for computing the volume of rainfall excess, 4) a unit hydrograph or kinematic overland flow model, 5) a time-of-concentration procedure, and 6) a channel routing method. The total watershed can then be subdivided so that the effects of local development within individual subwatersheds can be assessed for the entire watershed.

TABLE 9.3. Application of Flood Frequency Adjustment Method: Alhambra Wash Watershed, 1939–1951.

Water year (1)	% Imp. (2)	Annual max Q cfs (3)	Rank, m (4)	$\frac{m}{N+1}$ (5)	f_1 (6)	f_2 (7)	Adj. Q (8)	Adj. rank m (9)	$\frac{m}{N+1}$ (10)	f_1 (11)	f_2 (12)	Final adj Q, cfs (13)	Adj. rank m (14)
1939	26	4224	1	.071	1.39	1.61	4893	1	.071	1.39	1.61	4893	1
1940	26	2048	6	.429	1.55	1.84	2431	5	.357	1.52	1.81	2439	5
1941	27	3375	2	.143	1.42	1.69	4017	2	.143	1.42	1.69	4017	2
1942	29	1125	9	.643	1.65	1.92	1309	9	.643	1.65	1.92	1309	9
1943	30	2816	3	.214	1.52	1.75	3242	3	.214	1.52	1.75	3242	3
1944	32	2675	4	.286	1.59	1.77	2978	4	.286	1.59	1.77	2978	4
1945	33	2125	5	.357	1.63	1.82	2373	6	.429	1.67	1.84	2341	6
1946	34	1664	8	.571	1.72	1.91	1848	8	.571	1.72	1.91	1848	8
1947	36	1792	7	.500	1.73	1.87	1937	7	.500	1.73	1.87	1937	7
1948	39	1024	11	.786	1.90	2.01	1083	11	.786	1.90	2.01	1083	11
1949	42	1100	10	.714	1.93	1.98	1128	10	.714	1.93	1.98	1128	10
1950	43	900	13	.929	2.10	2.11	904	13	.929	2.10	2.11	904	13
1951	44	1024	11	.786	2.00	2.01	1029	12	.857	2.03	2.05	1034	12

There are numerous computer models available for urban stormwater analysis and design, each having its own input data structure. As an example, the SCS TR-20 program uses the 24-hour SCS design hyetograph, the SCS runoff equation, and the SCS curvilinear unit hydrograph (USDA SCS, 1969). An IDF curve is used to obtain the rainfall depth, which is used to dimension the ordinates of the SCS 24-hour dimensionless cumulative hyetograph. The SCS runoff equation is then used to compute the time distribution of rainfall excess. The incremental rainfall excess is then applied to the unit hydrograph in a convolution process to obtain the subarea design hydrograph. If necessary, the discharge hydrograph can be routed through channel reaches and combined with flood hydrographs from other subareas of the watershed. Urbanization is represented in the process by the runoff curve number (used to compute the rainfall excess) and by the time of concentration (used to dimension the unit hydrograph). It is noted that TR-20 is not particularly well-suited for storm sewer flow because of its inability to represent surcharge. While the details of the calculations vary, other computer models are also widely used to analyze urbanization effects on the flood hydrograph. Hydrologic models such as EPA SWMM and HEC-1 use the kinematic wave routing method for overland and channel/pipe flow. Like TR-20, HEC-1 is useful as an analysis tool but has limited capability to represent storm sewer flow. Selected models are described in a later section of this chapter.

II. PRECIPITATION IN THE URBAN WATERSHED

The management of urban storm runoff requires that the engineer/hydrologist conduct a careful study of historical rainfall data before selecting the design storm to be used in sizing the new project or in analyzing an existing one. It is presumed that a project will function properly if it can accommodate the design flow at full capacity. The project will fail to function as intended whenever the design event is exceeded. Some risk of failure must be accepted for economic reasons in selecting the magnitude of the design event. Thus, a storm sewer system designed to convey a 10-year storm runoff will surcharge and probably flood the street during a 25-year flood.

A return period, T, is defined as the average number of years between successive occurrences of a hydrologic event with a certain magnitude. The inverse of this return period is called the exceedance probability or the probability that this magnitude will be exceeded in any one year. Statistical flood frequency methods are used to analyze historical records for the return periods of different magnitude hydrologic events. The risk probability that the T-year event will occur at least once in N consecutive years is: $R = 1 - (1 - 1/T)^N$. The reliability probability is $1 - R$.

Because urban stormwater projects deal with storm runoff, the engineer ideally will choose an observed runoff event as the design event; however, urban runoff data is usually unavailable at most sites in the quantity and quality needed to obtain the peak discharge-return period and the volume-return period relations. Local rainfall data, on the other hand, are generally available. Thus, one approach to stormwater analysis and design is to transform a long-term rainfall record into a record of runoff before a design runoff event is identified. This can be achieved using a rainfall–runoff model, as discussed in a later section.

Another approach is to adopt national and regional rainfall intensity–duration–frequency (IDF) data found in various rainfall atlases. The U.S. Weather Bureau, now the National Weather Service, prepared isohyetal maps of rainfall depth for the entire United States for specified durations and return periods in a publication commonly known as TP 40 (Hershfield, 1961). These maps cover durations from 30 minutes to 24 hours and return periods from 1 to 100 years. TP 40 also includes interpolation procedures to determine rainfall depths for durations and return periods not included in the maps. Miller et al. (1973) updated the data in TP 40 and presented isohyetal maps for the 11 western states for 6- and 24-hour durations and return periods of 2 to 100 years in NOAA Atlas 2. It is noted that frequency analysis of local rainfall data may produce rainfall depths which differ significantly from NOAA Atlas 2. This is attributed to the regional averaging process in NOAA Atlas 2. The U.S. National Weather Service

publication known as HYDRO 35 (Frederick et al., 1977) is particularly useful to urban hydrologists because of the short duration rainfalls included. Many urban storm drainage projects deal with small drainage watersheds where storm durations less than 30 minutes need to be considered. HYDRO 35 includes isohyetal maps of rainfall depth for the eastern and central United States for 5-, 15-, and 60-minute durations and return periods of 2 and 100 years. HYDRO 40 provides comparable rainfall data for the southwest United States.

The purpose of this section is to describe how the available rainfall data can be used to identify a design runoff event. Basically, there are two fundamentally different approaches, namely the continuous simulation and the single-event design storm approaches. Both are discussed here. The single-event design-storm method is then elaborated further because of its simplicity and widespread use. Subsequent sections explain how to determine the elements of a design storm. The reader may refer to Chapter 2 for additional information on design storms.

A. Continuous Simulation and Single Events

In the continuous simulation approach, a chronological record of rainfall is used as input to a mathematical rainfall–runoff model of the drainage system being considered. The continuous rainfall record should be several decades long, usually at 1-hour time intervals, although the interval must be shorter for small drainage systems. The output from the mathematical model should be a chronological record of runoff. This record is then analyzed using statistical methods to determine the magnitude of the runoff that has a designated return period or exceedance probability. Since a moisture-accounting routine is incorporated in a continuous simulation, antecedent moisture conditions prior to rainfall events can be determined by the model itself after the initial start-up. Continuous simulation is particularly useful in detention projects in which the sequence of events and the inter-event time are important; however, the results of continuous simulation are meaningful only if an accurate rainfall–runoff model is used. The main disadvantages of continuous simulation are the large effort needed to prepare the input data, the long computer runs involved when hydrodynamic-based rainfall–runoff models are used, and the lack of short-duration rainfalls (15 minutes or less) required for small drainage systems.

In the single-event approach, a historical rainfall record is first analyzed to obtain the rainfall–return period relations. These are usually available in the form of intensity–duration–frequency or IDF curves. Then a single storm with return period equal to the design-life or maximum protection capacity of the project is chosen as the design storm. This event is used as input to a single-event rainfall–runoff model, along with an assumed antecedent condition for the urban watershed prior to the storm. The output is a design runoff hydrograph. An important assumption inherent in this approach is that the return period of the calculated runoff is the same as that of the design storm. This assumption is rarely met because the antecedent condition of the urban watershed, together with changes in the hydraulic behavior of the drainage system, can affect the runoff process significantly from one event to the next. Nevertheless, the assumption of equal rainfall and runoff frequency must be made in the design-storm approach.

A logical compromise between the design-storm and the continuous simulation approach is a pseudo-continuous method using a series of discrete major historical storms. In this approach, as summarized by Walesh (1989), a rainfall hyetograph and associated antecedent conditions are assembled for major storm events using a historical rainfall record of several decades. These hyetographs are input as multiple/discrete events to a rainfall–runoff model and the corresponding runoff hydrographs are produced. From the model output, the engineer/hydrologist then obtains the annual peak discharges and maximum volumes and these are analyzed statistically to develop discharge-probability and volume–probability relationships.

The continuous simulation method is data-intensive and time-consuming. Also, the mathematical representation of an urban watershed in a rainfall–runoff model may not be accurate enough to warrant continuous simulation. The design-storm approach, by comparison, is much simpler and is used extensively despite its limitations.

B. Elements of a Design Storm

A design storm is characterized in terms of a return period (or frequency), storm duration, average intensity (or total depth), spatial distribution, storm movement and direction, and time pattern of rainfall intensity (hyetograph). The selection of the design return period depends on several factors among which are the importance of the structure being designed, the project cost, the level of protection it will provide, and the damages that would result from failure of the structure. Typical design return periods vary from 2 to 25 years for storm sewers and 25 to 100 years for detention basins. Design return periods of 5 to 10, 10 to 25, and 25 to 50 years are common for culverts under highways carrying light, medium, and heavy traffic, respectively. Major highway bridges are designed to pass Q_{50} (the 50-year flood peak). Floodway delineation calculations, as presently administered by the Federal Emergency Management Administration (FEMA), are based upon Q_{100}. Design for erosion control in natural channels is usually based upon Q_2 or Q_5, since this flow range is regarded as dominant in forming stable channels. While street gutter design is commonly based on Q_2 or Q_5, it is obvious these appurtenances must be able to operate consistently with the main storm sewer system. Local drainage manuals and regulations must be used to determine the design return period for specific projects.

The design storm duration depends on the type of the project. For the design of hydraulic structures based on peak discharge, such as a storm sewer or culvert, the design storm duration should produce the largest peak discharge for a given return period. This will normally be the time of concentration for the drainage area above the design point. For flow detention projects, the duration causing the largest detention volume is the most critical. Generally, the critical storm duration can be determined only after trying several different storm durations and investigating the sensitivity of the peak discharge and the runoff volume to variable storm duration. For the design of storm sewers, the engineer/hydrologist should first estimate t_c, the time of concentration, and then obtain the average rainfall intensity from local intensity–duration–frequency relationships for a given return period and storm duration equal to t_c. These IDF relationships are discussed in a later section.

The spatial distribution of rainfall within an urban watershed can be highly nonuniform. The effect of watershed size on average areal rainfall is often accounted for by applying a reduction factor to the point rainfall values. Fig. 9.5 shows the percent reductions recommended by the U.S. Weather Bureau (1957) for the eastern United States. Most urban drainage catchments are sufficiently small that a rainfall reduction factor is not required. When large urban watersheds are investigated, the recommended reductions given in Fig. 9.5 can be employed. The U.S. Army Corps of Engineers has developed an equation for the reduction factors in Fig. 9.5. (USACE, 1990). Alternatively, the design storm can be assumed to occur only over part of the watershed with no areal reduction factor. Storm movement and direction of travel can have a significant impact on runoff peak, depending on the shape of the watershed. An elongated basin will experience higher runoff peaks when rainfall moves downstream from the headwater area rather than in the upstream direction. The location of the design storm center and the storm duration should be selected so that peak runoff rates and volumes will replicate real-world experience for a specified return period. Several locations must be tried before the most critical storm can be identified. The distribution of the total rainfall within the design storm can have a significant effect on the peak runoff rate. An early peaking storm may produce less runoff than a later-peaking one, having the same total rainfall. Several methods available to determine the time distribution of design rainfall are discussed in a later section. The interested reader also is referred to discussions by McPherson (1958), Preul and Papadakis (1973), Marsalek (1978), and Yen and Chow (1980; 1983).

C. Intensity–Duration–Frequency Relations

The relationships between average rainfall intensity, duration, and return period are usually presented as a graph as illustrated for Norfolk, Virginia in Fig. 9.6. These curves are obtained from local rainfall records using statistical frequency analysis techniques. Annual maximum or annual exceedance series of rainfall depths are extracted from historical records compiled by the National Weather Service or other

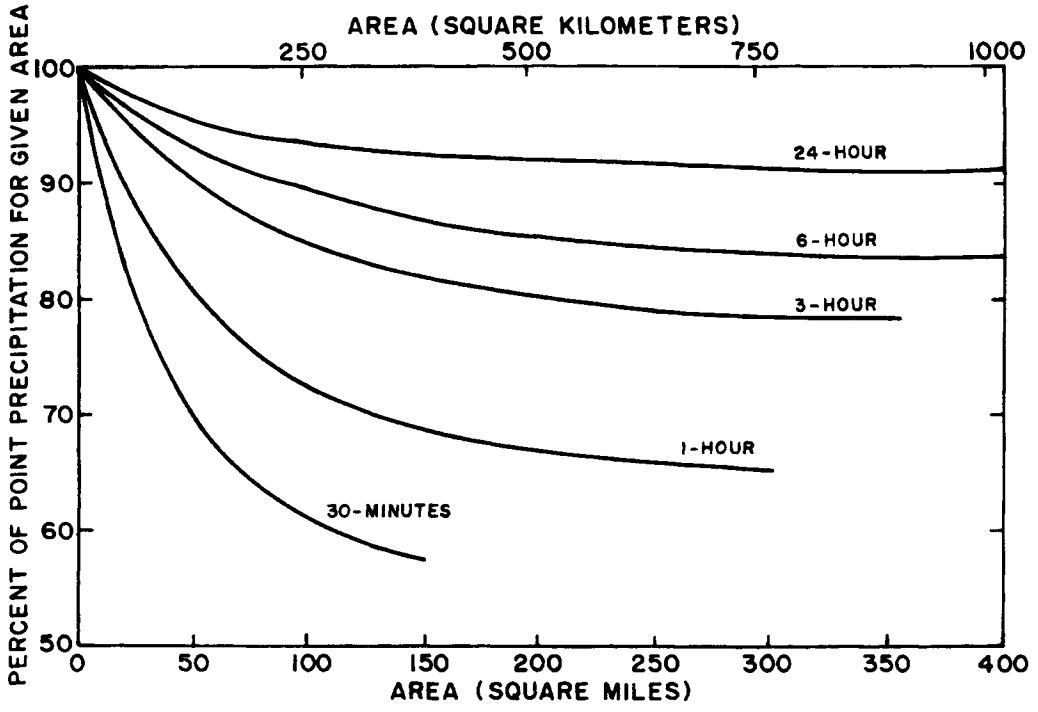


Figure 9.5.—Reductions in Point Rainfall to Obtain Areal Average Values (After Miller et al., 1973).

sources for the durations selected. Storm durations as small as 5 minutes are desirable for the urban drainage system. Then a frequency analysis is performed for each series. The Extreme Value Type I or the Gumbel distribution is commonly used for this purpose. The results are presented in chart form after dividing the depths by the respective durations to obtain the average intensities. These curves are commonly referred to as the intensity–duration–frequency (IDF) relations.

The IDF charts have already been developed for most major cities around the country, and they can be found in local drainage manuals and ordinances. Regional rainfall frequency maps reported in various publications, some of which were summarized earlier, can be used in the absence of local rainfall records or IDF relationships.

IDF relationships can also be expressed in equation form. Two common forms are

$$\bar{i} = \frac{A}{t_d^B + C} \tag{9.5}$$

and

$$\bar{i} = \frac{a}{(t_d + b)^c} \tag{9.6}$$

where $\bar{i}$ is the average intensity (cm/hr), t_d is the storm duration (minutes), and A, B, C, a, b, and c are fitting parameters. Wenzel (1982) presented the constants A, B, C, a, b, and c from a number of cities in the United States for storms of 10-year return periods. Table 9.4 shows these constants. Aron (1989) reported that for any return period, the constant b of Equation 9.6 is 15 and a and c can be determined from,

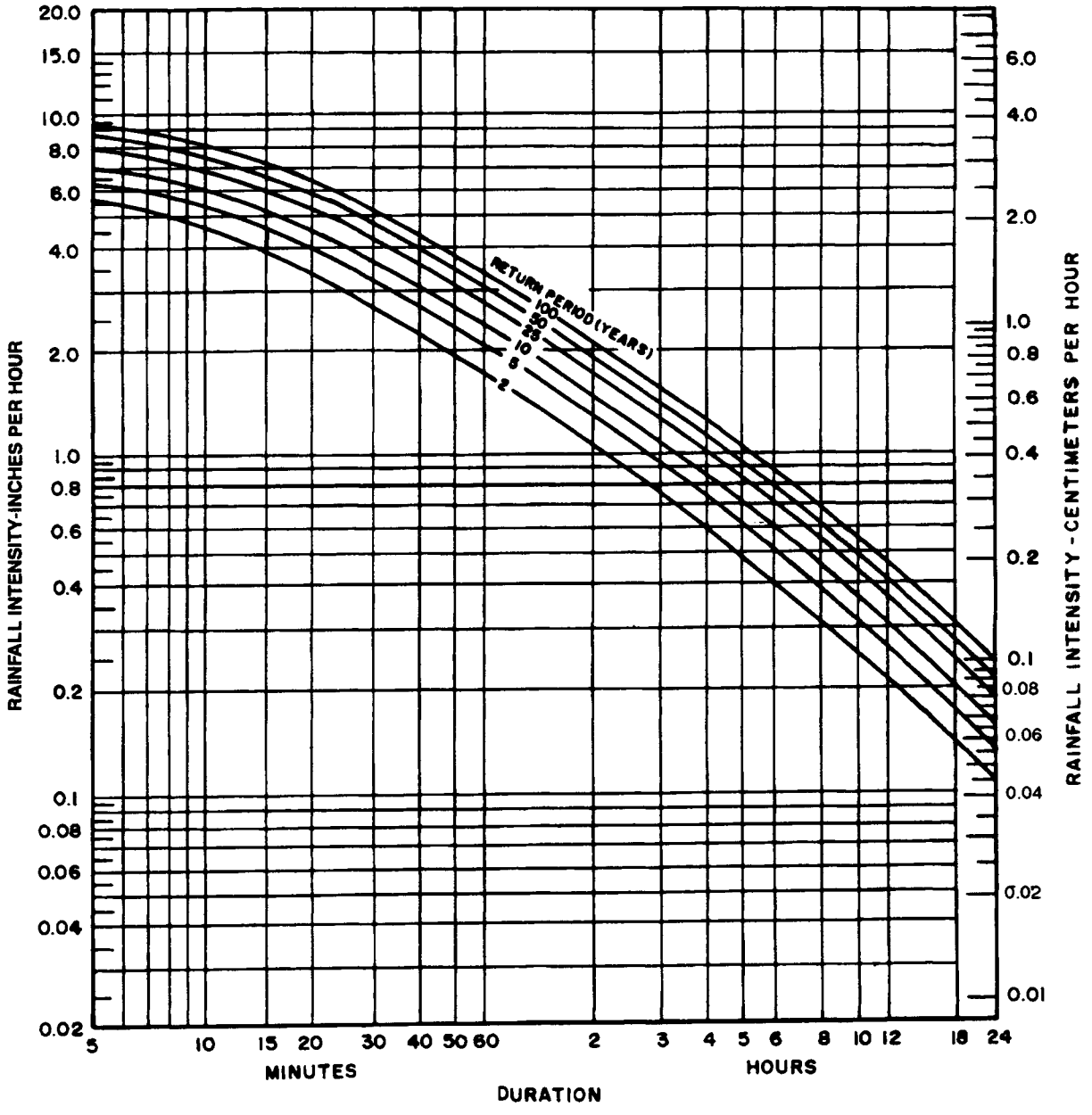


Figure 9.6.—Intensity-Duration Return Period Relationships for Norfolk, Virginia (VDOT, 1980).

$$a = 213.4 + 50.80\ (e^{0.394\, i_{60}}) \tag{9.7}$$

$$c = 0.2316\ ln\left(\frac{a}{i_{60}}\right) \tag{9.8}$$

where i_{60} is the 60 minute rainfall intensity in cm/hr for the return period being used. Aron's expressions were obtained by fitting Equation 9.6 to the generalized intensity–duration curves developed by Yarnell (1935). They provide a good fit to the Yarnell IDF curves for the range 5 min $\leq t_d \leq$ 120 min, but should not be extended beyond 120 minutes.

TABLE 9.4. Constants for IDF Relationships for 10-Year Return Period At Various Locations (After Wenzel, 1982).

Location	A	B	C	a	b	c
Atlanta	247.6	0.83	6.88	162.8	0.76	8.16
Chicago	241.0	0.88	9.04	154.7	0.81	9.56
Cleveland	187.2	0.86	8.25	120.9	0.79	8.86
Denver	245.4	0.97	13.90	129.0	0.84	10.50
Houston	247.4	0.77	4.80	249.7	0.80	9.30
Los Angeles	51.6	0.63	2.06	27.7	0.51	1.15
Miami	315.5	0.81	6.19	202.9	0.73	7.24
New York	198.4	0.82	6.57	130.6	0.75	7.85
St. Louis	265.9	0.89	9.44	154.9	0.78	8.96

Constants correspond to $\bar{i}$ in cm./hr. and t_d in min. in Eqs. 9.5 and 9.6.

D. Temporal Distribution of Rainfall and Design Storms

The time distribution of total rainfall within a storm is usually nonuniform. This distribution can have a significant effect on the computed peak runoff rate corresponding to a given design return period. Therefore, the time pattern of rainfall intensity needs to be specified for a complete description of a design storm. The uniform intensity distribution is the simplest form of the design storms, and it is associated with the Rational Method discussed in later sections. Other standard distributions employed extensively in urban storm drainage projects are nonuniform. Some of these distributions are discussed in the following subsections.

E. Soil Conservation Service Distributions

The U.S. Department of Agriculture, Soil Conservation Service (1986), developed four 24-hour rainfall distributions called Types I, IA, II, and III, respectively. These standard distributions represent different geographical locations within the United States as shown in Fig. 9.7. All design storms of shorter duration are included in these 24-hour rainfall hyetographs, and therefore these distributions are meant for both small and large watersheds. The four distributions are displayed in Fig. 9.8, and the fractions of the total rainfall are tabulated as a function of time in Table 9.5 in which P_T is total rainfall, and P is rainfall occurring between time zero and t.

An example 24-hour Type II hyetograph is developed in Table 9.6 for Norfolk, Virginia. A return period of 10 years is selected, and for this return period the average rainfall intensity for a 24-hour rainfall is found from the local IDF curves in Fig. 9.6, as 0.53 cm/hr (0.21 in./hr). The total rainfall depth, P_T, is 0.53 cm/hr. × 24 hours = 12.72 cm (5.04 in). The peak portion of the calculated hyetograph is displayed in Fig. 9.9.

F. Other Design Storm Hyetographs

Various other rainfall distributions have been proposed in the literature. Hershfield (1962) reported time–distribution relationships based on hourly rainfall data from 50 widely separated stations with different rainfall regimes. Durations of 6, 12, 18, and 24 hours were considered. An average relationship was obtained for these four durations as shown in Fig. 9.10. Hershfield (1962) points out that such an average curve gives no indication of the distribution from an individual storm. Therefore, a rearrangement of the rainfall distribution for a particular event is not unreasonable. An example 10-year, 2-hour storm hyetograph is developed for Norfolk, Virginia in Table 9.7 assuming the average Hershfield distribution can be used for 2-hour durations also. The total rainfall depth of P_T = 7.47 cm (2.99 in) can be obtained from Fig. 9.6 as described in the preceding paragraph. The resulting design storm hyetograph is displayed in Fig. 9.9. Other balanced design storm distributions are described by the U.S. Army Corps of Engineers (1982).

Huff (1967) presented a thorough analysis of large storms that occurred in east central Illinois between 1955 and 1966. The storm durations varied from 3 to 48 hours. The storms were grouped according to the

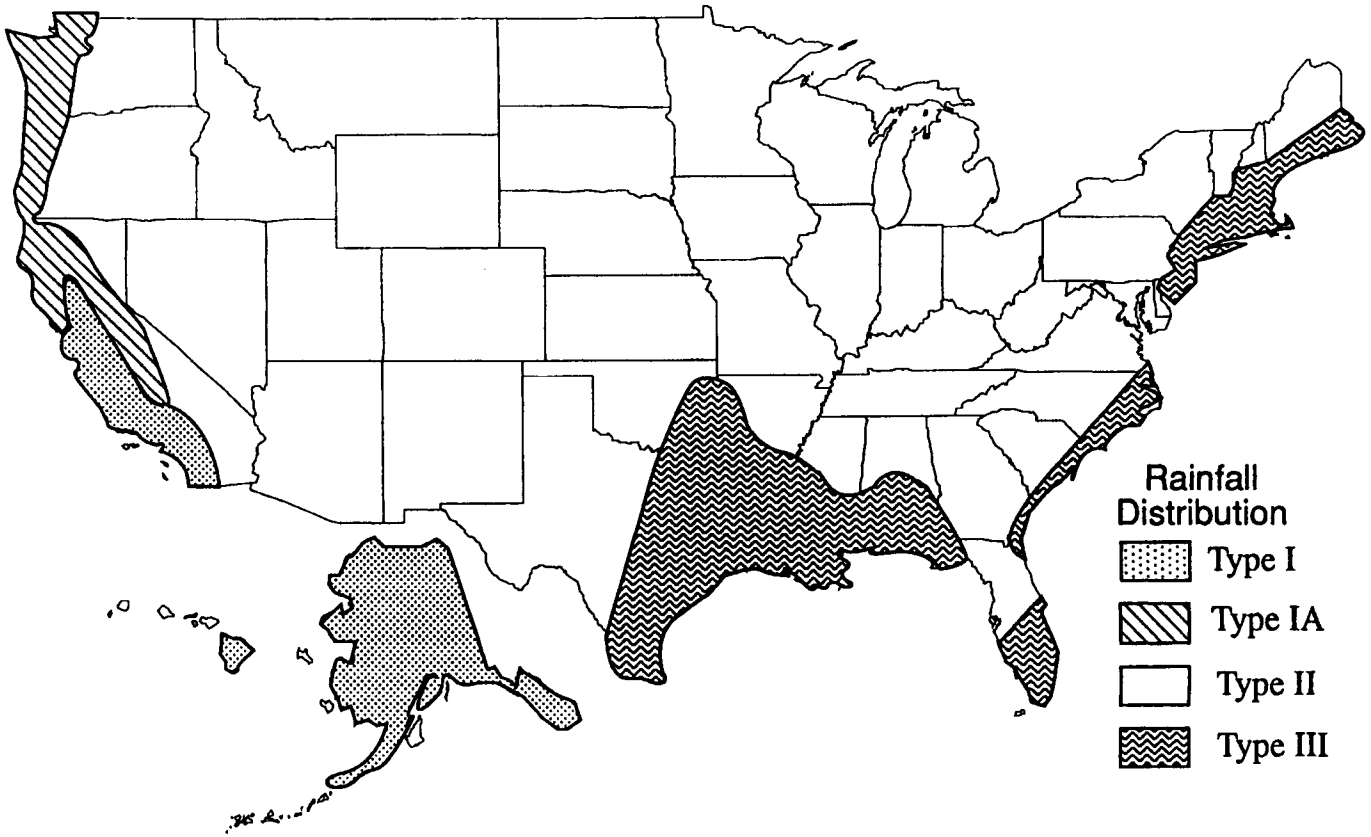


Figure 9.7.—Approximate Geographical Boundaries for SCS 24-hr Rainfall Distributions (After USDA SCS, 1986).

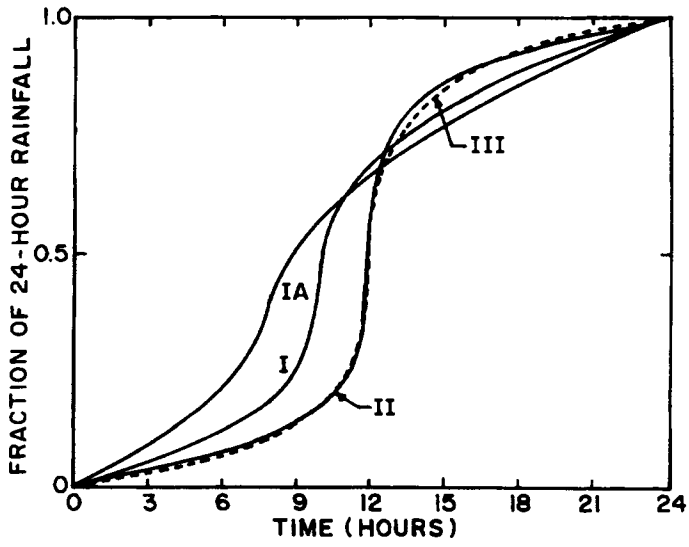


Figure 9.8.—Soil Conservation Service 24-hr Rainfall Distributions (After USDA SCS, 1986).

quartile in which rainfall was the heaviest. Then probability distributions were obtained for the storms in each quartile. For small areas and short durations, first-quartile storms were the most frequent. The median distribution for all the four quartiles is shown in Fig. 9.10. An example second quartile Huff hyetograph is developed in Table 9.8 for Norfolk using a return period of 10 years and a duration of 2 hours. As before, the total rainfall is P_T = 7.47 cm. The resulting hyetograph is plotted in Fig. 9.9.

In Fig. 9.9 it is clear that there are major discrepancies in the three rainfall distributions. Since these are all statistically derived, there is no way to assign a measure of absolute accuracy to any of the distribu-

TABLE 9.5. Soil Conservation Service Rainfall Distributions.

	Fraction of Total Rainfall, P/P_T			
Hour t	Type I	Type IA	Type II	Type III
0	0	0	0	0
2.00	0.035	0.050	0.022	0.020
4.00	0.076	0.116	0.048	0.043
6.00	0.125	0.206	0.080	0.072
7.00	0.156	0.268	0.098	0.089
8.00	0.194	0.425	0.120	0.115
8.50	0.219	0.480	0.133	0.130
9.00	0.254	0.520	0.147	0.148
9.50	0.303	0.550	0.163	0.167
9.75	0.362	0.564	0.172	0.178
10.00	0.515	0.577	0.181	0.189
10.50	0.583	0.601	0.204	0.216
11.00	0.624	0.624	0.235	0.250
11.50	0.654	0.645	0.283	0.298
11.75	0.669	0.655	0.357	0.339
12.00	0.682	0.664	0.663	0.500
12.50	0.706	0.683	0.735	0.702
13.00	0.727	0.701	0.772	0.751
13.50	0.748	0.719	0.799	0.785
14.00	0.767	0.736	0.820	0.811
16.00	0.830	0.800	0.880	0.886
20.00	0.926	0.906	0.952	0.957
24.00	1.000	1.000	1.000	1.000

TABLE 9.6. Example SCS 24-hr Type II Hyetograph for Norfolk, VA.

t (hrs)	P/P_T	P (cm)	Δp (cm)	Δt (hrs)	i (cm/hr)
0	0	0			
			0.611	4.00	0.153
4.00	0.048	0.611			
			0.915	4.00	0.229
8.00	0.120	1.526			
			0.344	1.00	0.344
9.00	0.147	1.870			
			0.203	0.50	0.406
9.50	0.163	2.073			
			0.229	0.50	0.458
10.00	0.181	2.302			
			0.293	0.50	0.586
10.50	0.204	2.595			
			0.394	0.50	0.788
11.00	0.235	2.989			
			0.611	0.50	1.222
11.50	0.283	3.600			
			0.941	0.25	3.764
11.75	0.357	4.541			
			3.892	0.25	15.568
12.00	0.663	8.433			
			0.916	0.50	1.832
12.50	0.735	9.349			
			0.471	0.50	0.942
13.00	0.772	9.820			
			0.343	0.50	0.686
13.50	0.799	10.163			
			0.267	0.50	0.534
14.00	0.820	10.430			
			0.764	2.00	0.382
16.00	0.880	11.194			
			0.915	4.00	0.229
20.00	0.952	12.109			
			0.611	4.00	0.153
24.00	1.00	12.720			

tions. Rather, it is left to the engineer/hydrologist to select a rainfall distribution and duration that will produce a conservative design event for the particular project at hand. As a minimum, the rainfall duration should be twice the time of travel through the drainage area to allow development of a full hydrograph. The use of a 24-hour storm is considered highly conservative in the case of local stormwater detention projects. This may be readily justified depending on the downstream damage area.

III. HYDROLOGIC LOSSES IN DEVELOPING WATERSHEDS

Hydrologic losses refer to processes which act on rain or snow before it turns into rainfall excess, snowmelt, and surface runoff. These processes generally include evaporation of water from the land and vegetative leaf surface, interception of rainfall by vegetation, depression storage on the land surface (paved or unpaved) and infiltration of water into the soil matrix. If the effects of a single storm event of short duration are of primary interest, evaporation losses can be neglected; however, it has been shown by Horton (1933) that during a sequence of low-intensity rainfalls, evaporation can enhance considerably the interception capacity of a dense leaf canopy.

Interception, as part of surface retention, varies depending on the type of vegetation and maturity and extent of the canopy cover. Depression storage, which is the major part of surface retention in urban

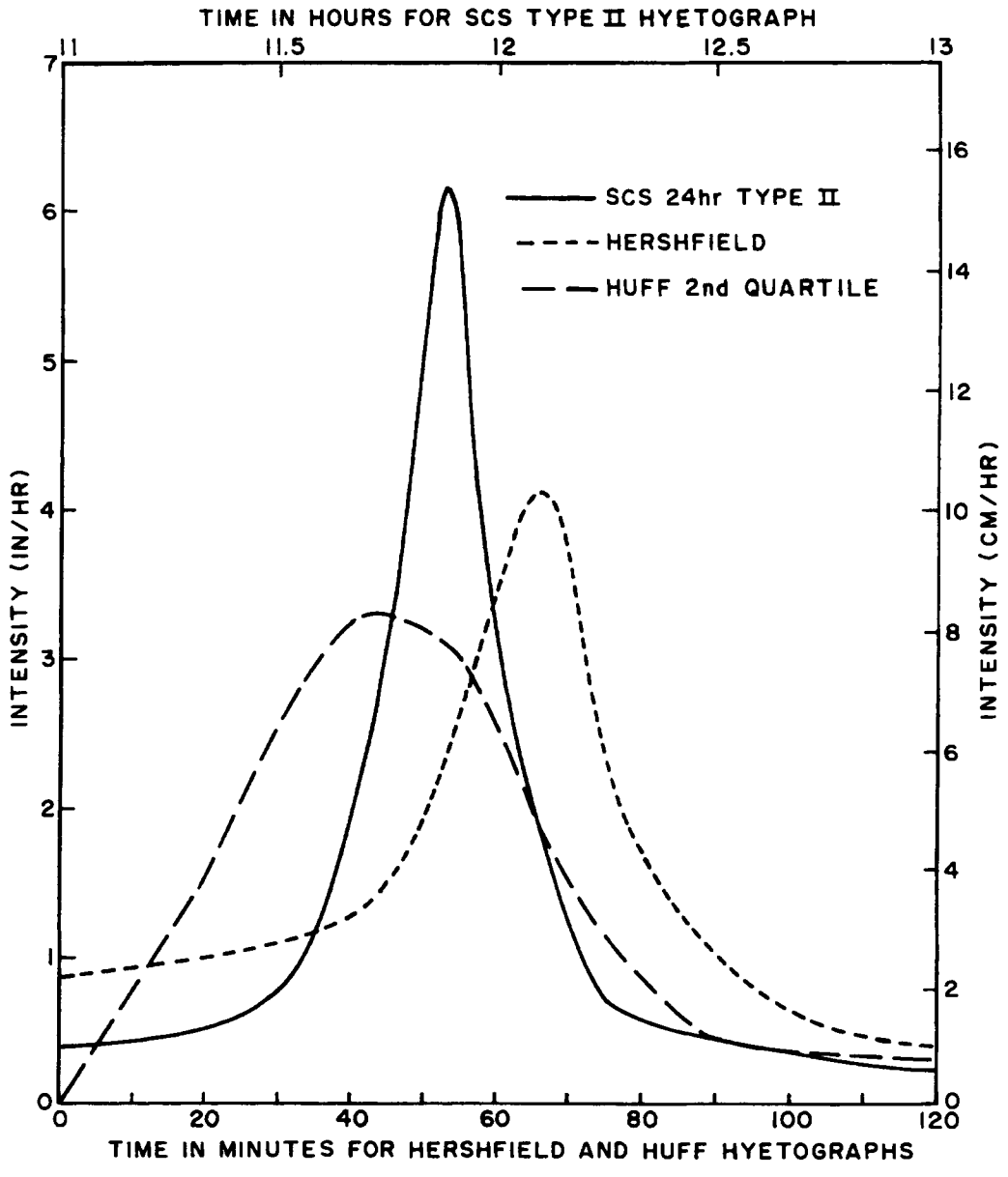


Figure 9.9.—Comparison of SCS Type II, Hershfield, and Huff Second Quartile Hyetographs.

basins, depends on the irregularities in the surface topography. Finally, infiltration comprises the major portion of abstractions from pervious land surfaces. Like interception, infiltration tends to be strongly enhanced by warm temperatures and a dense vegetation.

A. Interception

References to interception can be found in most hydrology textbooks and some professional journal papers, but almost all of these refer to Horton's (1919) observations taken in his laboratory and published in a highly detailed 20-page article. Horton concluded that evaporation can play a major role in the interception process, if the total leaf surface of the projected area is dense and the storm endures at least several hours. For these conditions, Horton proposed the equation:

$$J = S_j + k_j E D_s \tag{9.9}$$

TABLE 9.7. Example Hershfield Hyetograph for Norfolk, VA.

t (min)	t/t_d	P/P_T	P (cm)	ΔP (cm)	ΔP/Δt (cm/hr)
0	0	0	0		
				0.448	2.24
12	0.1	0.06	0.448		
				0.486	2.43
24	0.2	0.125	0.934		
				0.560	2.80
36	0.3	0.20	1.494		
				0.672	3.36
48	0.4	0.29	2.166		
				1.196	5.98
60	0.5	0.45	3.362		
				2.091	10.46
72	0.6	0.73	5.453		
				0.971	4.86
84	0.7	0.86	6.424		
				0.523	2.62
96	0.8	0.93	6.947		
				0.299	1.50
108	0.9	0.97	7.246		
				0.224	1.12
120	1.0	1.0	7.470		

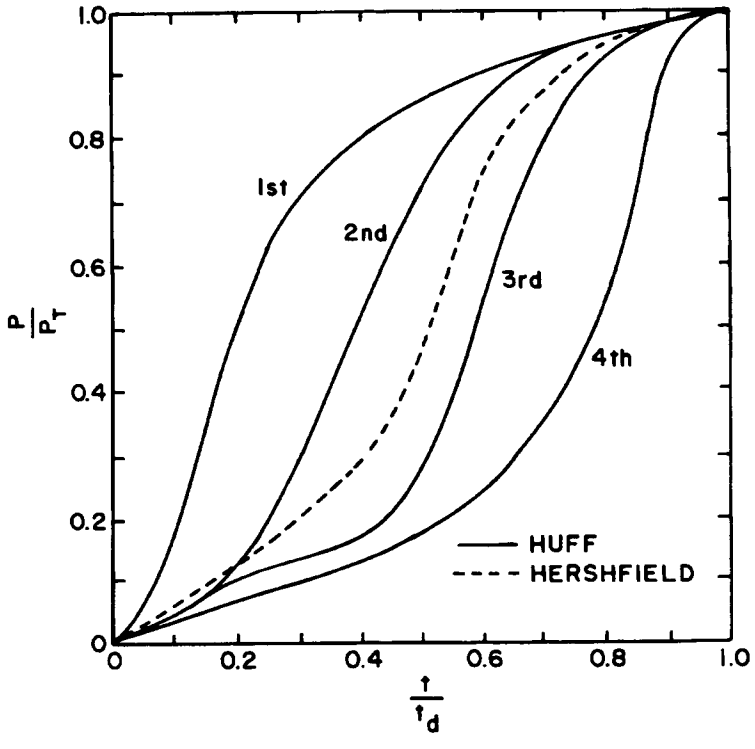


Figure 9.10.—Hershfield and Huff Rainfall Distribution.

TABLE 9.8. Example Second Quartile Huff Hyetograph for Norfolk, VA.

t (min)	$\frac{t}{t_d}$	$\frac{P}{P_t}$	P (cm)	ΔP (cm)	$\frac{\Delta P}{\Delta t}$ (cm/hr)
0	0	0.0	0		
				0.247	1.24
12	0.1	0.033	0.247		
				0.687	3.44
24	0.2	0.125	0.934		
				1.307	6.54
36	0.3	0.300	2.241		
				1.681	8.41
48	0.4	0.525	3.922		
				1.568	7.84
60	0.5	0.735	5.490		
				0.934	4.67
72	0.6	0.860	6.424		
				0.486	2.43
84	0.7	0.925	6.910		
				0.231	1.16
96	0.8	0.956	7.141		
				0.195	0.98
108	0.9	0.982	7.336		
				0.134	0.67
120	1.0	1.000	7.470		

in which J is the interception (cm), S_j is the interception storage capacity (cm), on the projected area, k_j is the ratio of the evaporation surface to the projected area, E is the evaporation rate (cm per hour), and D_s is the storm duration (hours).

A modification of the Horton equation was later proposed, reflecting the notion that not all of the interception capacity is immediately filled. The modified equation reads:

$$J = S_j \left[1 - \exp\left(-\frac{P}{S_j}\right)\right] + k_j \; E \; D_s \tag{9.10}$$

in which P is cumulative precipitation (cm).

For situations in which evaporation is negligible or difficult to evaluate, Horton proposed the alternative equation:

$$J = a + b\,P^n \tag{9.11}$$

in which a, b, and n are parameters to be found experimentally. Horton's parameter values for selected tree types and for short crops are shown in Table 9.9 (for J and P in inches).

Two problems can be found with Equation 9.11 and the parameter values. In the first place, interception can exceed precipitation for small values of P. Secondly, the suggested value of the exponent n is 1.0 for most tree types, which implies that interception continues to increase at an undiminished rate as precipitation increases. Because of these deficiencies, Aron (1992) analyzed a set of data collected from summer storms by Horton (1919). Equation 9.12 was developed by linear regression on the logarithmic values of P and J.

$$J = c\,P^m \tag{9.12}$$

Chapter 10

WATER WAVES

I. INTRODUCTION

Surface water waves are a mechanism for transferring energy from a source across the water surface to a shoreline or structure. Common sources of energy are the wind, the gravitational pull of the moon and sun, seismic activity, or a moving vessel. These waves range in period from a few seconds for wind-generated waves up to twelve and twenty-four hours for the major components of the tide. They are commonly the most important design factor for shoreline stabilization works, marinas housing moored vessels, and navigation control structures.

The Water Waves chapter consists of eleven related sections. The level of presentation in these sections varies. For example, the sections considered less important to hydrologists and hydraulic engineers are kept brief while others significant to their practice are more comprehensive. The presentation is a first-level treatment in that the state of the art is reviewed and pertinent references are given for detailed analyses. Example problems are included where necessary to illustrate applications of the theories and principles.

The sections in the chapter are separated for an orderly presentation of several complex subjects in water waves. The chapter begins with a review of basic water wave theory and the calculation of wave properties. Wind waves are discussed in the following section, including wave generation, prediction, and statistical properties in deep and shallow waters. Ship-generated waves are briefly described in the next section. The effects of waves on structures is reviewed in the section titled Wave–Structure Interaction. Wave–current interaction is outlined in the following section. The longer period phenomena, including water level fluctuations, astronomical tides, storm generated surges, tsunamis, and basin resonances, are described in the next three sections. Water surface probability analysis follows. The selection of design waves and water levels and water surface elevation probability analysis is reviewed in the last section of the chapter.

It is noted that each of these topics is quite broad in scope, and it is virtually impossible to present in this handbook all aspects of water waves of greatest interest to hydrologists and hydrologic engineers. Appropriate references are included and listed at the end of the chapter to provide more details on the subject of water waves. The Shore Protection Manual (SPM) (USACERC, 1984), is a good overall reference.

II. WAVE THEORY

A. Fundamentals and Classification

Wave motions can be calculated using wave theories, which have been developed since 1845. These theories are simplified versions of nature, often involving such assumptions as horizontal bottoms or

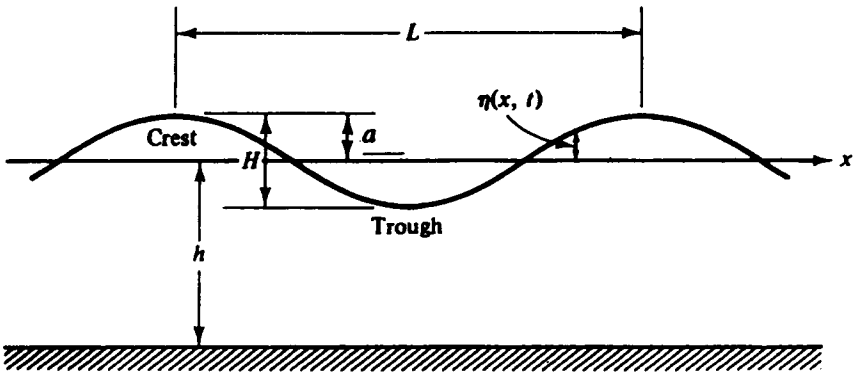


Figure 10.1.—Schematic of a Water Wave.

inviscid flow. There are a variety of these wave theories that are useful for different situations; for example, there are shallow water theories and deep water theories and there are small amplitude wave theories and large amplitude theories. This section will discuss the background of some of these theories and illustrate their uses.

Almost all wave theories have a common theoretical basis, such as water is incompressible and fluid motions are irrotational. Further the bottom is assumed to be flat and impermeable. The equations governing these wave theories are either the Laplace equation or Euler's equations of motion.

Important wave parameters include height, H, which is the vertical distance from the wave crest to the wave trough (as illustrated in Fig. 10.1), wave length, L, which is the distance between two crests, and the period, T, the time for two waves to pass a point. (L and T are later demonstrated to be related.) Another important parameter is water depth, denoted by d hereafter, except in Figs. 10.1, 10.2, and 10.3, where it is denoted as h. Dimensional analysis shows that the following dimensionless ratios are important:

$$H/L, d/L \tag{10.1}$$

The first dimensionless number is the wave steepness, and the second is the relative water depth. Steeper waves tend to be more forceful and also require more elaborate theories to describe them. The relative water depth is used to determine whether the wave is in deep or shallow water, requiring a different theory. Dividing these two dimensionless numbers yields a third number, H/d, which often occurs in wave breaking calculations.

B. Linear (Airy) Wave Theory

The simplest and surprisingly most versatile wave theory is the Airy theory. It is also called linear theory due to the mathematical linearization of the nonlinear wave problem. A full description of the linear wave theory is provided by Dean and Dalrymple (1991).

The water surface displacement of this wave has the following trigonometric form:

$$\eta(x,t) = \frac{H}{2}\cos[2\pi(x/L - t/T)] = \frac{H}{2}\cos\left(\frac{2\pi}{L}(x - Ct)\right) \tag{10.2}$$

at location x and at time t. The presence of the 2π in the argument of the cosine function ensures that the wave repeats itself in a distance equal to a wave length and at a time equal to a wave period. The speed of the wave in the positive x direction is C, which is the distance the wave travels in one wave period divided by the wave period, or C = L/T.

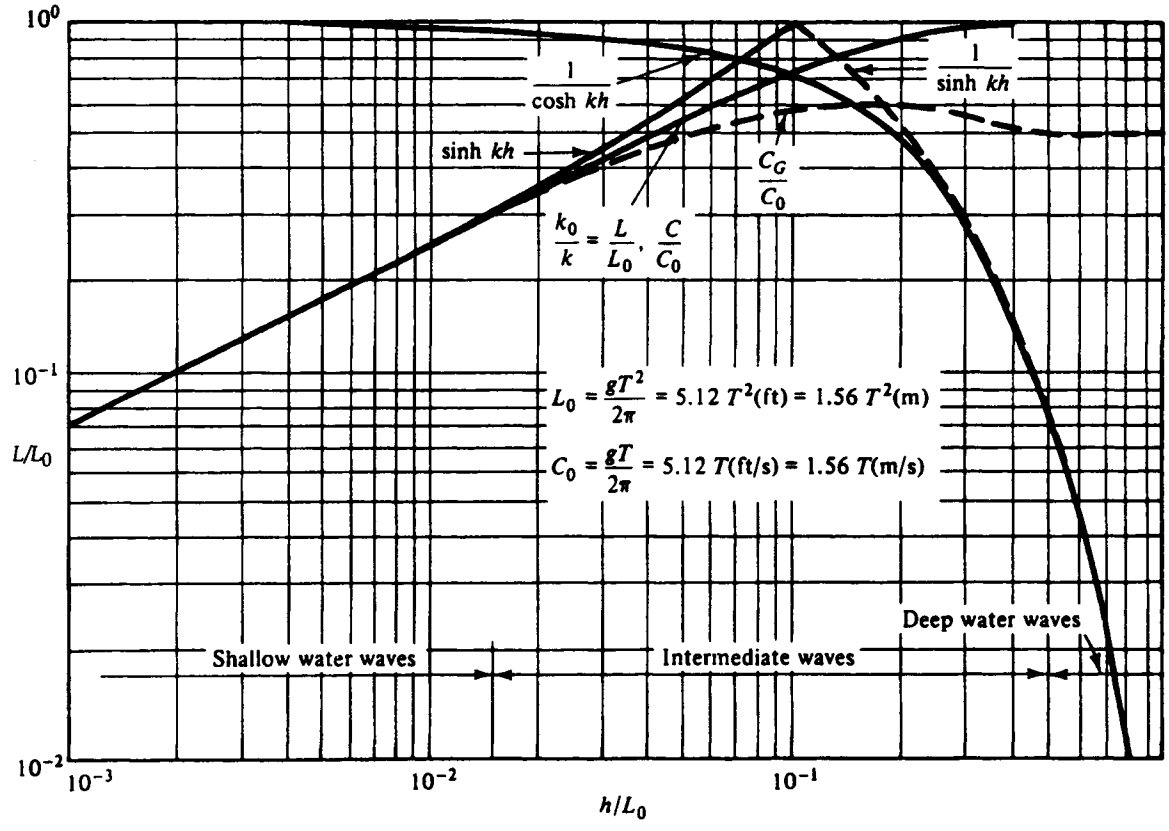


Figure 10.2.—Graphical Means to Calculate Wave Number, Wave Speed, and Group Velocity (Dean and Dalrymple, 1991).

For linear theory, the wave length and wave period are related through the (frequency) dispersion relationship

$$L = \frac{gT^2}{2\pi}\tanh\left(\frac{2\pi d}{L}\right) \tag{10.3}$$

where g is the acceleration of gravity. This equation involves the wave length in the argument of the hyperbolic tangent function as well, and therefore the equation must be solved numerically (often the Newton–Raphson method is used) or graphically. Fig. 10.2 provides a means to determine the wave number.

Note that as the relative water depth decreases, the wave length decreases as well. For deep or shallow water, the equation can be solved directly. For waves in deep water, the relative water depth is large, $d/L > 1/2$. In this case, the hyperbolic tangent has a value approaching unity, and the above equation becomes

$$L_o = \frac{g}{2\pi}T^2. \tag{10.4}$$

Example The wave period is known to be 5 seconds and the water depth is 3.05 m (10 ft.). What is the wave length, L?

First, the deep water wave length is calculated: $L_o = g\,T^2/2\pi = 9.81\,(5)^2/\,(2\times3.1415\ldots) = 39$ m. Therefore $d/L_o = 0.078$. Using Fig. 10.2, $L/L_o = 0.65$. This means that the local wave length is $L = 0.65\,(39) = 25.4$ m, which is very close to the exact value of 25.1 m (about a 1% error in reading the graph).

In shallow water, the argument of the hyperbolic tangent is small, $d/L < 1/20$, such that the function approaches its argument in value, and the dispersion relationship becomes

$$L_s = \sqrt{gd}\,T. \tag{10.5}$$

An interesting result for shallow water is the wave speed, C_s ($= L_s/T$) is simply $\sqrt{gd}$, independent of the wave period.

As another alternative to the graphical method (or as an initial guess for the Newton–Raphson solution of Equation 10.3), an accurate approximation to the dispersion relationship is that of Fenton and McKee (1989), which has a maximum error less than 1.7%, well within most design criteria. Their relationship is

$$L = \frac{g}{2\pi}T^2\left(\tanh\left[(2\pi\sqrt{d/gT})^{3/2}\right]\right)^{2/3}. \tag{10.6}$$

For the previous example problem, this formula gives L = 25.48 m, which differs only from the exact solution by 1.5%.

The rate at which energy is carried by the waves is different from the speed of the wave form, C. The speed at which energy is transported is the group velocity, C_g, and it is related to the wave speed. In deep water, $C_g = C/2$, while in shallow water, $C_g = C$. The ratio of the two is

$$n = \frac{C_g}{C} = \frac{1}{2}\left(1 + \frac{\frac{4\pi d}{L}}{\sinh\frac{4\pi d}{L}}\right). \tag{10.7}$$

Fig. 10.2 can be used to calculate the group velocity conveniently. For the previous case, where $d/L_o = 0.078$, the value of $C_g/C_o = 0.56$. Since $C_o = g\,T/(2\pi) = 7.8$ m/s, $C_g = 4.1$ m/s (about a 3.5% error).

The particle velocities under the Airy wave depend on the particle's position relative to the wave crest and the bottom. The horizontal and vertical velocities are

$$u(x,z,t) = \frac{H\pi}{T}\frac{\cosh\frac{2\pi}{L}(d+z)}{\sinh\frac{2\pi}{L}d}\cos\left(\frac{2\pi}{L}(x - Ct)\right) \tag{10.8}$$

$$w(x,z,t) = \frac{H\pi}{T}\frac{\sinh\frac{2\pi}{L}(d+z)}{\sinh\frac{2\pi}{L}d}\sin\left(\frac{2\pi}{L}(x - Ct)\right) \tag{10.9}$$

where z is measured upwards from the water surface. The velocities are maximum at the water surface, u is in-phase with the wave form, and w goes to zero at the bottom ($z = -d$) to satisfy the impermeability condition.

The pressure under the wave is composed on the hydrostatic term due to the presence of the water and a dynamic term due to the presence of the wave

$$p(x,z,t) = -\rho g\left(z - \eta(x,t)\frac{\cosh\frac{2\pi}{L}(d+z)}{\cosh\frac{2\pi}{L}d}\right) \tag{10.10}$$

C. Nonlinear Wave Theories

Mathematically, the linear theory is valid for infinitesimally small steepness, that is, for $(2\pi/L)a << 1$. In practice, most waves have greater steepness and should be treated with a more sophisticated wave theory to correctly model wave characteristics. For example, in linear theory, the wave length does not depend on the wave height, while for nonlinear theories higher waves go faster than smaller waves.

The choice of a correct theory depends on the water depth of interest. For deeper water, the Stokes theories are often used, while for shallow water the Boussinesq theories are most often used.

D. Stokes Theory

The Stokes theories come about through a perturbation expansion in wave steepness. This means the steeper the waves, the more terms must be added to the original linear theory to provide a correct representation of the wave motion.

As an example, the second order Stokes theory provides a water surface displacement of the following form:

$$\eta(x,t) = \frac{H}{2} cos2\pi(x/L - t/T) - \frac{H^2\pi \cosh(2\pi d/L)}{8L \sinh^3(2\pi d/L)} (2 + \cosh(4\pi d/L)) cos4\pi(x/L - t/T). \quad (10.11)$$

The added term oscillates twice as fast as the linear term adding to the linear term at the crest and trough, causing the crests to steepen and the troughs to flatten. There is also a change in the mean water level, referred to as the set-down.

$$\overline{\eta}(x,t) = - \frac{\pi H^2}{4L \sinh(4\pi d/L)} \quad (10.12)$$

The set-down is negligible in deep water and becomes more pronounced in shallow water. The fifth order Stokes theory adds three more contributions to the expression of the wave form. Fenton (1985) provides a description of the fifth order Stokes theory.

Higher order theories are solved at the cost of considerable mathematics. It is far easier today to use a numerical procedure for solving these theories. For example, Dean (1965) developed the Stream Function wave theory, which allows for very high order theory that is valid in both deep and shallow water. Dean (1974) has tabulated the results of 40 different wave cases, spanning the range from deep to shallow water for a variety of wave heights (up to breaking). These tables can be used by interpolation to provide accurate predictions of wave characteristics, velocities, and accelerations. Dalrymple (1974b) developed a version of the Stream Function theory for the case of currents that converged on wave height.

E. Shallow Water Theories

In shallow water, where the water depth to wave length ratio is very small, an alternative theoretical formulation provides a more compact expression of the wave field. This is the so-called Boussinesq theory, which has two well-known formulations, the solitary wave and the cnoidal wave theories.

F. Solitary Wave Theory

The solitary wave consists solely of a wave crest and has an infinite length. Russell (1838) was the first to discuss this type of wave motion, which he observed in canals where horsedrawn canal boats were being used. The free surface of the solitary wave is described by

$$\eta(x,t) = a\ sech^2\left(\sqrt{\frac{3a}{4d}}\frac{x}{d}\right). \tag{10.13}$$

The volume of water contained in the wave crest is

$$V = 4d\sqrt{\frac{ad}{3}}. \tag{10.14}$$

Most of this wave volume (95%) is contained in a finite distance, 2 ℓ, about the wave crest, where

$$\ell = \frac{2.12d}{\sqrt{a/d}}. \tag{10.15}$$

The speed of the solitary wave is approximately

$$C = \sqrt{gd}\left(1 + \frac{a}{2d}\right). \tag{10.16}$$

G. Cnoidal Wave Theory

The cnoidal wave theory provides for a periodic wave train in shallow water. These waves, described by Korteweg and deVries (1895), are characterized by a steepness factor, K. The wave form is

$$\eta = a\ cn^2\left(\sqrt{\frac{3a}{4dK^2}}\frac{x}{d}\right). \tag{10.17}$$

The cn function is a Jacobean elliptic function with modulus, K. Wiegel (1964) shows the variations of the wave form with K. The mathematical complexities of the elliptic functions usually result in the use of numerical models in shallow water design.

H. Numerical Models

The wave-induced motion in shallow water is described by the Boussinesq equations for variable depth (Peregrine, 1967). Solutions of these equations are obtained by a variety of methods—finite differences (Abbot et al., 1978), spectral means (Freilich and Guza, 1984), and parabolic approaches for small angle of wave incidence (Liu et al., 1985). Computational models are now available commercially that describe the motion of shallow water waves over irregular bathymetry, around structures, and in harbors.

I. Shoaling/Refraction

Waves travel slower in shallow water than in deep water due to the decrease in wave length (see Equation 10.3). If a long-crested wave encounters variable-depth water, the portions of the wave crest in deeper water will travel faster than those in shallower water, resulting in changes in the local wave direction, called refraction. For simple bathymetry, the wave direction change can be computed from the wave speed with Snell's Law

$$\frac{\sin\theta_1}{C_1} = \frac{\sin\theta_2}{C_2} \tag{10.18}$$

where the subscripts denote the wave angle and wave speed at one location compared to those on the other side of a change of depth. The wave angle is measured from a line perpendicular to a bottom contour. Following a section of wave crest as the wave propagates towards shore allows the construction of wave rays, which are lines that are always parallel to the wave direction. Graphical techniques exist to compute the wave rays by hand (JSPM) (USACERC, 1984); however, this is easier done by computer.

The change in wave height due to changes in wave length and direction between two points, denoted by 1 and 2, is obtained by

$$H_2 = H_1\sqrt{\frac{C_{g,1}}{C_{g,2}}}\sqrt{\frac{b_1}{b_2}} = K_s K_r \tag{10.19}$$

where K_s is the shoaling coefficient, involving the ratio of the group velocities at the two locations and K_r is the refraction coefficient, which is a function of the ray spacing, b.

For the simplest case of straight and parallel contours in the alongshore direction, Fig. 10.3 provides a means to calculate wave height and direction change at a given depth as a function of the deep water wave height and direction and a dimensionless ratio, $d/(gT^2)$.

Example A wave with a deep water wave height of 1 m and a wave period of 8 seconds encounters a shoreline characterized by straight and parallel contours. The deep water wave angle of incidence (measured from the shoreline perpendicular) is 30°. Find the wave height and direction in 2.5 m water depth.

> For this case, $d/(gT^2) = 2.5/(9.81 \times 8^2) = 0.004$. Entering Fig. 10.3 from the right where $\theta_o = 30°$, move horizontally to the vertical line corresponding to $d/(gT^2) = 0.004$. Interpolating between the solid lines corresponding to θ, the wave direction in shallow water is found to be $\theta = 11°$ and $K_rK_s = 1.10$, which gives H (1.10) = 2.2m.

For more complicated bathymetry, computer programs exist for refraction and shoaling based either on ray tracing algorithms (Noda, 1974) or rectangular grid methods (Dalrymple, 1988).

J. Diffraction

Diffraction is the change in wave direction effected by strong variations in wave height. For example, waves passing a breakwater have very low wave heights behind the breakwater. Diffraction causes a wave direction change to allow energy to flow to the low wave height regions. Penney and Price (1952) showed how the analogy to optics can be used for wave calculations. Graphic examples of waves passing very long breakwater sections or through gaps in breakwaters, all for the case of constant depth, are available (USACERC, 1984). The important variables are gap spacing and the wave length.

For variable depth in the vicinity of the breakwaters, both refraction and diffraction must be computed. The combined effects are conveniently described by the mild-slope equation, developed by Berkhoff (1972). Numerous models of the mild slope have been developed.

Parabolic modelling of the mild-slope equation (Radder, 1979) has led to the development of a class of models that permit the rapid calculation of wave fields where both refraction and diffraction is important (Kirby and Dalrymple, 1983, 1984; Liu and Tsay, 1983).

K. Wave Breaking

In very shallow water, waves break as the wave length decreases monotonically with depth and the height increases until finally the wave is unstable and breaks. The simplest criterion for wave breaking is that the wave height is a significant proportion of the water depth, $H = \kappa d$, where κ is the breaker

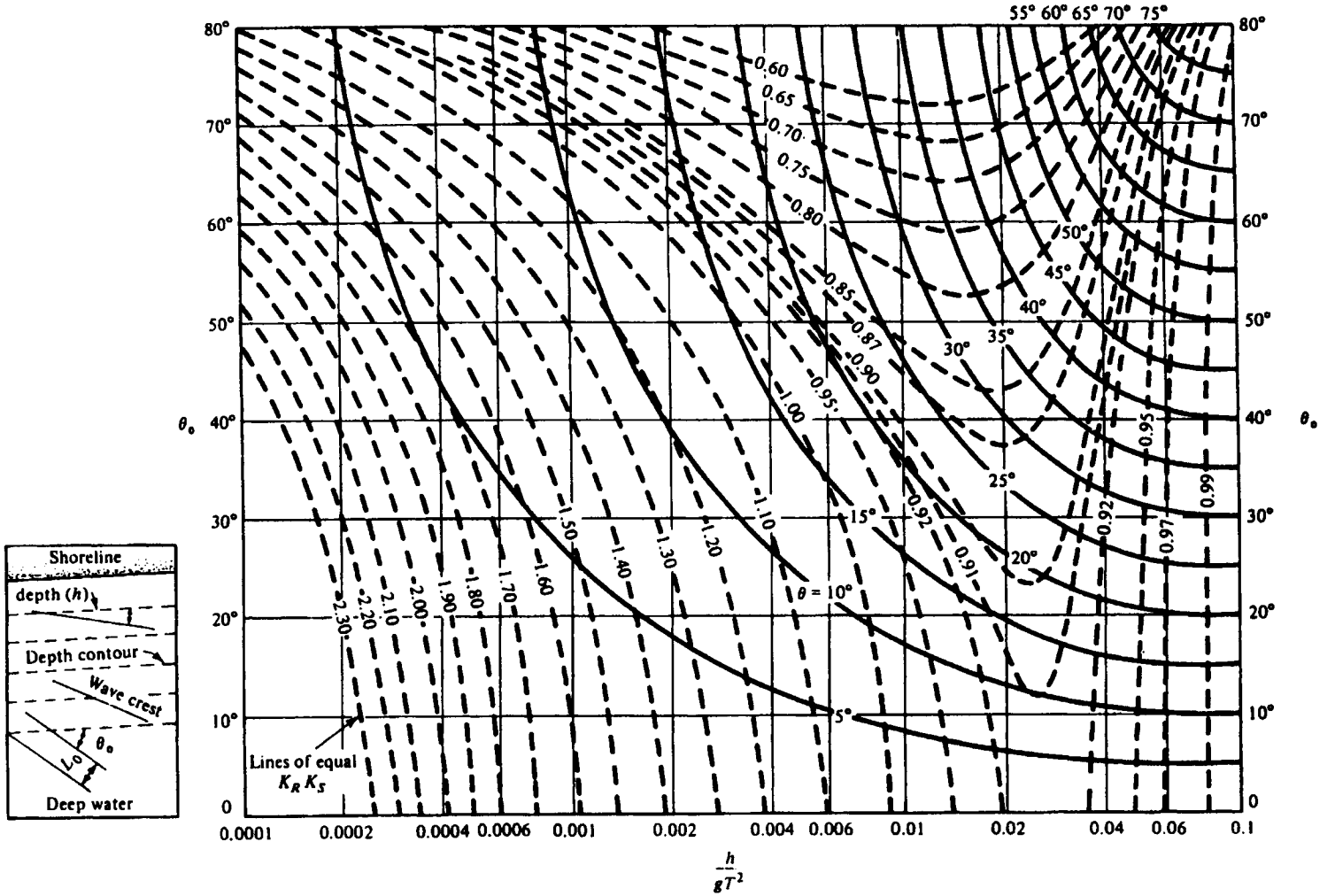


Figure 10.3.—Changes in Wave Direction and Height Caused by Refraction on a Shoreline with Straight and Parallel Bottom Contours (USACERC, 1984).

index that has a range of $0.78 < \kappa < 1.5$. The particular value for κ depends on the slope of the bottom and the wave steepness, H/L. For mild slopes, the smaller values apply. This implies that over a flat bottom, the maximum wave height is 0.78 times the water depth.

Battjes (1974b) introduced the surf similarity parameter, ζ, to describe a number of nearshore phenomena, including wave breaking. This parameter is defined as:

$$\zeta = \frac{m}{\sqrt{H/L_o}} \tag{10.20}$$

where m is the local beach slope. The surf similarity parameter has a typical range of $0.1 < \zeta < 5.0$, ranging from very mild slopes to steep slopes. In Table 10.1, the breaking index is shown as a function of the surf similarity parameter.

Waves break in a variety of ways. On mild slopes spilling breakers occur where the foaming crest of the wave slides down the wave front. Plunging breakers (surfers' delight) occurring on steeper beaches have crests that fall forward of the wave front in a big jet, often trapping air. Surging breakers occur on very steep beaches where breaking is dominated by wave reflection.

For a vertical wall, the waves do not break but are totally reflected, leading to standing waves in front of the wall, with a doubling of wave height.

Example Determine the breaking depth for a 1.5-m high wave on a beach that has a 1:10 slope. The wave period is 6 seconds.

> The beach slope m = 0.1. $L_o = 1.56(6)^2 = 56.16$ m (from Equation 10.4). The surf similarity ζ in Equation 10.20 is $0.10/\sqrt{1.5/56.16} = 0.61$. From Table 10.1, $\kappa \sim 1.0$. Therefore, the breaking depth would be 1.5 m.

More sophisticated wave models take into account the details of wave breaking (Dally et al., 1985) or the spectral nature of waves (Battjes and Janssen, 1978; Thornton and Guza, 1983).

III. WIND WAVES

A. Description of Irregular Waves

Contrary to the assumptions of wave theory presented in the preceding section, wind waves in nature are a statistical process. The height and period of individual waves passing a fixed point in space during a stationary sea state vary considerably and more or less randomly with time. Useful statistical descriptions and parameters of a real sea state have been developed so that wave information can be conveniently used in engineering applications. The descriptions and parameters are discussed in this section. General references on wind waves and more complete and detailed information is given elsewhere (Dean and Dalrymple, 1991; Goda, 1985; IAHR, 1986; Ochi, 1982; Sorensen, 1978; USACE, 1989).

A closely related engineering concern is the long-term description of the ensemble of sea states that can occur at a site. The long-term description includes wave climate and extreme wave analysis. These topics are addressed in the section titled Selection of Design Water Waves and Levels.

TABLE 10.1. Surf Similarity Parameter and Breaker Index and Reflection Coefficient.

$\zeta = \frac{m}{\sqrt{H/L_o}}$	0.1	0.5	1.0	2.0	3.0	4.0
Breaker Type	Spilling		Plunging		Surging	No Breaking
κ	0.8	1.0	1.1	1.2		
R	0.001	0.01	0.1	0.4	0.8	

The concept of a "significant wave height" and a "significant wave period" that can be used to characterize a sea state is appealingly simple. It suggests a direct transition from the experimental results in a laboratory wave tank and regular wave theory to the phenomena that occur in nature. The concept of a significant wave height and period was first introduced when sailors were asked to report the height and period of the larger, well-formed waves, and omit entirely the low and poorly-formed waves as part of the synoptic weather reports from ships. Comparisons of early wave gage records with observations led to the conclusion that the wave height given by observers was approximately equal to the average height of the one-third highest individual waves, $H_{1/3}$. The parameter $H_{1/3}$ is referred to as "significant wave height," and the corresponding period is the "significant wave period," $T_{1/3}$. Thus, formally, significant height is the average height of the one-third highest individual waves in a record, and significant period is the average period of the one-third highest individual waves in a record. In practice, significant wave height and period can be estimated by several approaches, as discussed in the following paragraphs.

B. Wave Measurements

The primary information from any wave observation technique is generally an estimate of significant wave height and period. Many techniques provide additional valuable data. Commonly-used techniques are described in the remainder of this section. Their particular advantages and disadvantages are given in Table 10.2.

Wave staff observations are obtained with a gage that pierces the water surface. In its simplest form, the wave staff is a vertical rod with visible marks on it at measured intervals. It allows an observer to obtain better quantitative estimates of wave characteristics by watching the undulating water motions at the staff; however, most wave staffs operate as electrical sensors. Additional information about wave staffs is available from Parker (1982).

In addition to the advantages and disadvantages of the wave staff listed in Table 10.2, the staff is often inexpensive relative to other instrumental observation techniques. The visually-observed staff is especially economical, and it can be very useful in low-budget projects. In low energy environments, the staff gage can be mounted on a spar buoy as an alternative to a rigid mounting structure.

Pressure cell observations are obtained with a pressure-sensitive gage mounted under the water surface. Pressure cells can be situated anywhere in the water column as long as they are below the elevation of the lowest expected wave trough at the lowest expected tide level. Pressure cells are often placed on a small tripod which rests submerged on the bottom. The pressure cell senses dynamic pressure fluctuations created by passing surface waves. The magnitude of the fluctuations decreases exponentially with distance below the surface. The pressure fluctuations are converted to an electronic

TABLE 10.2. Wave Observation Techniques (USACE, 1989).

Technique	Water Depth[1]	Advantages[2]	Disadvantages[3]
Shore observations	3	1	1
Staff gage	1,2,3	2,5	2,3
Pressure gage (shore-connected by cable or telemetry)	2,3	3,5	3,4,5
Pressure gage (internally recording)	2,3	3,4	3,4,5,6,7
Accelerometer buoy	1,2	4,5	5,8,9

[1]Water Depths: 1 = deep, 2 = intermediate, 3 = shallow

[2]Advantages: 1 = inexpensive; 2 = direct measurement of surface waves; 3 = relatively reliable for unattended use; 4 = relatively simple installation; 5 = adaptable to real time monitoring.

[3]Disadvantages: 1 = low accuracy; 2 = sturdy support structure required for exposed sites; 3 = subject to fouling by marine growth; 4 = divers needed for installation and maintenance; 5 = indirect measurement of desired surface waves; 6 = prone to loss due to mechanical failure (leaking) or damage by bottom trawlers; inaccurate positioning and loss of marker buoys; 7 = frequent maintenance required for power supply and recorder; 8 = prone to loss due to collision or failed mooring; 9 = relatively expensive.

signal which can be recorded at the gage or sent to a shore station by either an armored electrical cable or a surface buoy transmitter. Additional information about pressure cells is available (Forristall, 1982).

Accelerometer buoy observations are obtained with a surface-following buoy which is usually spherical in form. Buoys are routinely moored in water depths of less than 200 m. Buoys sense vertical acceleration, which is usually integrated twice electronically to give a record of surface displacements. The observations may be recorded in the buoy or transmitted to a shore station either directly or via a satellite link. Some buoys are equipped with additional motion sensors so that wave direction information may be obtained.

A variety of other wave recording instruments are available for practical use although those discussed in the preceding subsections are most widely used. The instruments are categorized as in-situ instruments and remote sensing devices. A variety of ground-based and airborne remote sensing devices are available but most are not suited for routine data collection.

In-situ instruments include acoustic, ultrasonic, optical (laser), and radar instruments, all of which transmit a signal toward the water surface from above or below. The reflected signal is then received and interpreted. Another approach to in-situ data recording consists of operating several instruments together at a site and analyzing the records jointly to get additional information, particularly wave direction. Typical combinations are a pressure cell or staff with two orthogonal horizontal current meters, and a spatial array of staffs or pressure cells. More information on these in-situ instruments is available (Parker, 1982; Seymour, 1982).

C. Wave Analysis and Statistics

Wave records are usually collected as a digital time series of surface elevation or subsurface pressure. Typical steps in analyzing a digital wave record are schematized in Fig. 10.4 and briefly discussed in this section. The editing step should include a check for waves at short and long periods outside the range of wind wave periods, which is 1 to 30 seconds in the ocean. Apparent sources of waves at undesired periods are tides, water level oscillations, surf beats, electronic drift, electronic noise, and transmission

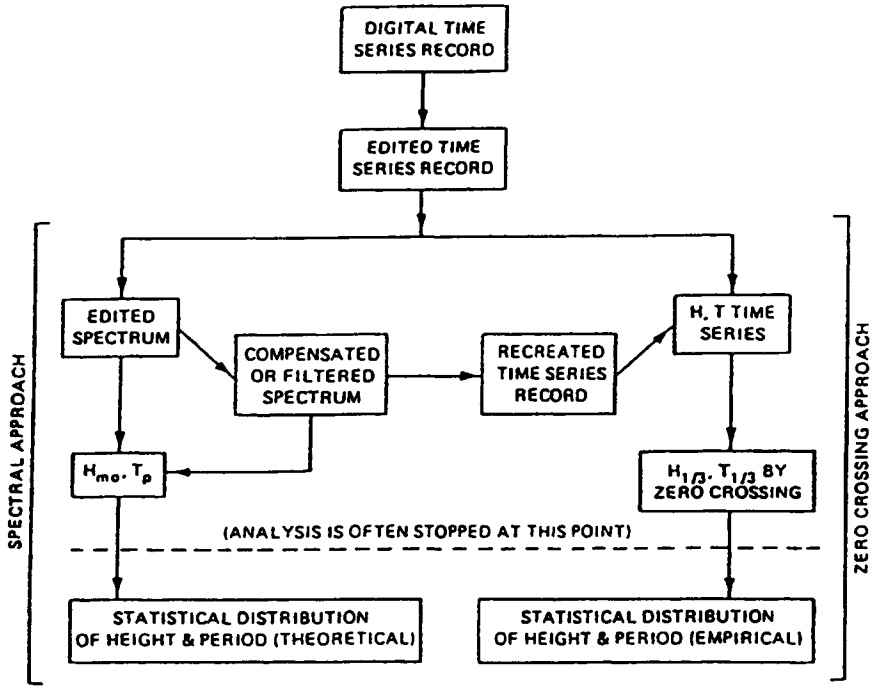


Figure 10.4.—Analysis of a Digital Wave Record (USACE, 1989).

interference. If any of these apparent waves are significant in the record, they may distort estimates of wind wave characteristics. Short- and long-period contamination can be identified visually or by numerical tests. It can be removed by filtering or by considering the spectrum discussed in the following subsections.

A sea state can be analyzed by extracting information from the time series or by transforming the time series into a frequency spectrum and extracting information from the spectrum (spectral analysis). Although spectral analysis is most common, both approaches are widely used and give complementary information.

Significant height and period may be estimated directly from the time series by identifying all individual waves in the record. The procedure that is most widely used and most easily applied on a digital computer is known as the zero-crossing method. Each wave in the record is identified as an event between two successive points at which the wave trace crosses the mean in a downward moving direction (Fig. 10.5). The time between crossings is the wave period, T. Wave height, H, is defined as the elevation difference between the highest point (crest) and lowest point (trough) of each wave. Significant wave height is computed as the average height of the highest one-third zero-crossing waves. Other wave height statistics, such as the root-mean-square wave height, are also easily computed. Significant period is computed as the average period of the one-third highest waves. This procedure is called the zero downcrossing analysis procedure because each wave is defined by two downcrossings of the mean.

By the spectral approach, significant height is estimated from the spectral density integrated over the range of wind wave frequencies (IAHR, 1986). A standardized analysis package for wind waves should be used when the spectral approach is desired. A comprehensive package has the advantages of automatically removing nonwind wave energy from the results and options for compensating pressure or acceleration signals to give estimates of surface waves. The significant height estimate is sometimes referred to as H_{m0} to clearly identify that it was obtained by a spectral approach. Wave period estimated by the spectral approach is the period corresponding to the highest energy density in the spectrum. It is called significant or peak period, T_p. In relating spectral parameters to traditional time series parameters, it is generally assumed that

$$H_{mO} = H_{1/3}. \tag{10.21}$$

This relationship is reasonable as long as the wave shapes are not severely deformed by shallow water depth or high wave steepness. In very shallow water, an approximation may be used (Hughes and Borgman, 1987). The accepted approximate relationship for wave period estimates is

$$T_p = 1.05\, T_{1/3}. \tag{10.22}$$

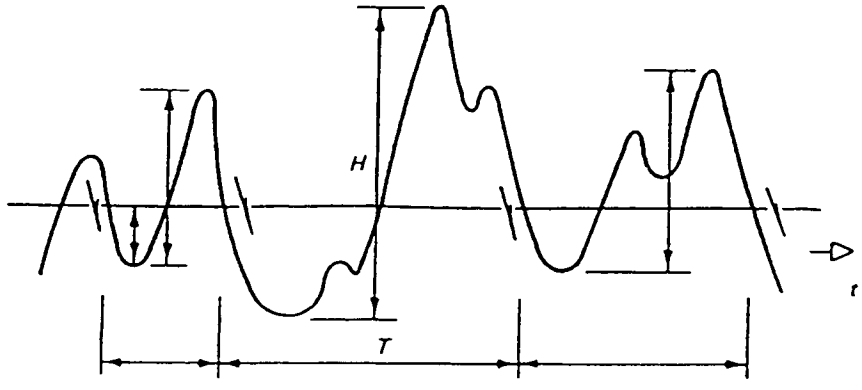


Figure 10.5.—Zero-Downcrossing Waves.

This relationship may not be realistic at exposed ocean locations where several wave trains may arrive simultaneously from different generation areas. Wave direction is an additional parameter that is often of considerable engineering importance. Wave direction can be estimated from measurements only through spectral analysis. The term H_s is commonly used to designate a generalized significant wave height. When it is used, the method for estimating it and whether it represents $H_{1/3}$ or H_{m0} should be clarified.

The distribution of individual wave heights in a sea state is almost always well-approximated by the Rayleigh distribution function. The Rayleigh distribution is given by

$$P(H \geq \hat{H}) = e^{-\left(\frac{\hat{H}}{H_{rms}}\right)^2} \tag{10.23}$$

where $P(H \geq \hat{H})$ = probability that a wave in the distribution will exceed height $\hat{H}$, $\hat{H}$ = given wave height, H_{rms} = root-mean-square wave height, and e = 2.7183. Other statistical wave height parameters can be easily related to H_{rms} through the Rayleigh distribution. The more commonly used parameters are listed in Table 10.3.

An estimate of the maximum wave height in a sea state can also be obtained from the Rayleigh distribution as

$$H_{max} = H_{rms}\sqrt{\ln N} = 0.71\, H_{1/3}\sqrt{\ln N} \tag{10.24}$$

where H_{max} is the maximum individual wave height, ln is the logarithm with respect to the base e, and N is the number of waves occurring. The parameter N can be estimated for a storm as the duration of the intense part of the storm divided by the peak or significant wave period. The distribution of individual wave periods is much more variable than the distribution of individual wave heights. It is also of less engineering importance than the height distribution.

D. Wind Parameters and Fetch

Simplified methods have been developed for estimating wave parameters during wind wave generation. The simplified methods are particularly useful when quick, low-cost estimates are needed. They provide accurate results for areas with relatively simple wind conditions, fetches, and water depths. The methods presented here are taken from the SPM (USACERC, 1984).

Estimation of the wind speed is a critical, and often difficult, first step. Winds over the wave growth area are needed for estimation. Winds can be estimated from measurements or from surface pressure information on synoptic weather charts. Wind speed must be adjusted to be consistent with the wave growth equations. That is, the wind speed should be adjusted for elevation, averaging time, air–water temperature differences, local land effects (if winds are from a measurement on or near land), and nonconstant coefficient of drag. Details of the adjustments are given in the SPM (USACERC, 1984). At a minimum, the elevation and drag coefficient adjustments should be made. The elevation adjustment can be approximated by the following equation if the observation elevation is less than 20 m:

$$U_{10m} = U_z\left(\frac{10}{z}\right)^{1/7} \tag{10.25}$$

TABLE 10.3. Wave Height Relationships Based on the Rayleigh Distribution.

Average of 1% highest waves = 2.36 H_{rms} = 1.67 $H_{1/3}$
Average of 10% highest waves = 1.80 H_{rms} = 1.27 $H_{1/3}$
Average of 1/3 highest waves = 1.42 H_{rms} = 1.00 $H_{1/3}$
Average of all waves = 0.89 H_{rms} = 0.63 $H_{1/3}$

where U_{10m} is the wind speed at 10 m elevation (needed for wave prediction), U_z is the wind speed observed at elevation z, and z is the elevation of the wind observation, in meters. The drag coefficient correction is:

$$U_A = 0.71\ U^{1.23} \tag{10.26}$$

where U_A is the wind-stress factor (adjusted windspeed) and U is the wind speed after other adjustments have been made, such as elevation, etc., in meters per second.

Estimates of the wind duration may also be needed for wave prediction if the fetch is relatively long. The duration usually must be estimated by judgment and interpolation using available information. A fetch is defined as the length of an overwater region in which the wind speed and direction are reasonably constant. A fetch should be defined such that wind direction variations do not exceed 15 degrees and wind speed variations do not exceed 2.5 m/s from the mean. When the point of interest is located downwind from a coastline, the coastline limits the fetch. In inland waters (bays, rivers, lakes, and reservoirs), fetches are limited by landforms surrounding the water body. Shorelines are usually irregular. In such cases, fetch can be defined as a radial average over an arc of 24 degrees centered on the wind direction with radials placed at 3-degree intervals (Fig. 10.6).

E. Deep Water Wave Prediction

Significant wave height and peak period generated by local winds can be estimated from the wind speed (corrected as prescribed in the SPM, USACERC, 1984), duration, and fetch. The estimation is based

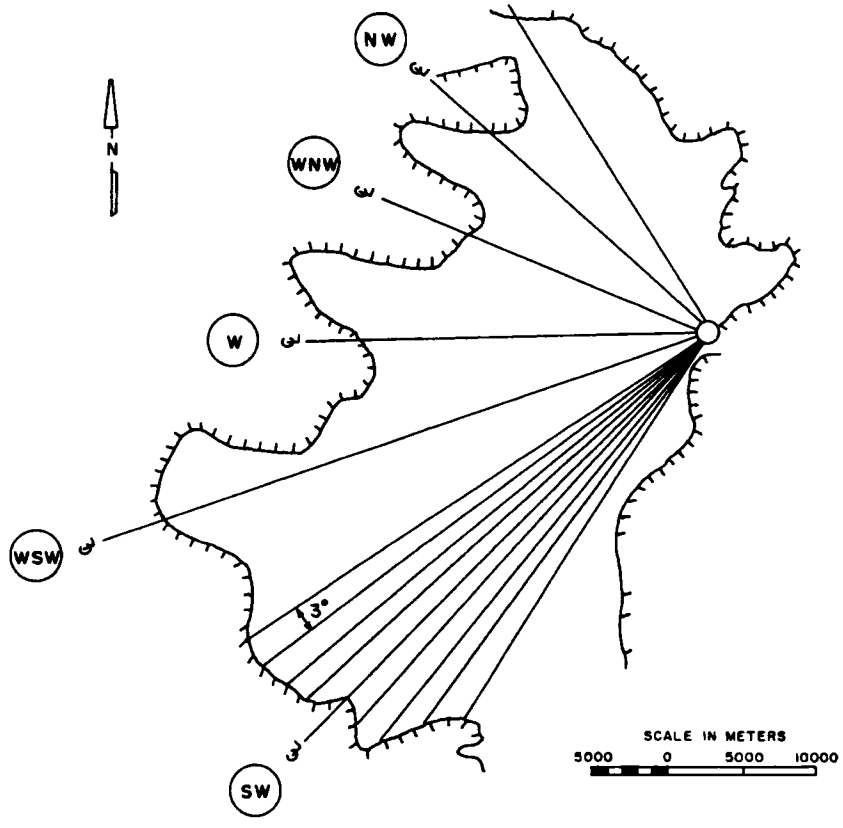


Figure 10.6.—Fetch Estimation on Small Water Bodies with Irregular Shape (USACE, 1989).

on equations developed from the Joint North Sea Wave Project (JONSWAP) experiment. The equations are:

$$H_{mO} = 0.01616\, U_A\, F^{1/2} \tag{10.27}$$

$$T_p = 0.6238\, (U_A\, F)^{1/3} \tag{10.28}$$

where F is the fetch (in km), provided H_{m0} is in meters, T_p in seconds, and U_A in meters per second.

On lakes and reservoirs, wave conditions are almost always limited by the wind speed and fetch. On the ocean, where fetches can be very long, the time duration over which the wind blows often limits wave growth more than the fetch. The minimum duration which will not limit wave growth is approximated as:

$$t = 0.893 \left(\frac{F^2}{U_A}\right)^{1/3} \tag{10.29}$$

where t is duration in hours. Special procedures for use with hurricanes and other tropical storms are available (USACERC, 1984).

F. Shallow Water Wave Growth

If the predominant depth of water over the fetch is less than one-half the deepwater wavelength, wave growth is affected by the bottom. When shallow water wave growth occurs, the known wind speed, duration and fetch can be used to predict significant wave height and period from the following equation:

$$H_{1/3} = 0.283 \left(\frac{U_A^2}{g}\right) \tanh\left[0.530 \left(\frac{gd}{U_A^2}\right)^{3/4}\right] \tanh\left[\frac{0.00565 \left(\frac{gF}{U_A^2}\right)^{1/2}}{\tanh\left[0.530 \left(\frac{gd}{U_A^2}\right)^{3/4}\right]}\right] \tag{10.30}$$

where d is water depth, usually determined as a rough average of the depths in the downstream half of the fetch, g is acceleration due to gravity = 9.80 m/sec^2, and tanh is the hyperbolic tangent function. The minimum duration required to reach fetch-limited wave growth is given by:

$$t = 0.0537 \left(\frac{U_A}{g}\right) \left(\frac{gT}{U_A}\right)^{7/3}. \tag{10.31}$$

Dimensional units in the above equations must be consistent with the units used for gravity.

G. Computer Modeling

Because of the continually increasing capabilities and availability of digital computers, comprehensive numerical wave models are becoming essential tools for practical engineering work. Models vary greatly in complexity. The simplest wave growth model is a computer program to solve the wave growth equations presented above. The most complex models are comprehensive spectral models which operate on a grid and simulate natural wave processes including growth, dissipation, propagation, and nonlinear

interactions between different wave frequencies. The models can be used for both wave hindcasting (wind input based on data from past events) and wave forecasting (wind input based on predictions for the future). A microcomputer software package for coastal engineering applications, called the Automated Coastal Engineering System (ACES), is available (Leenknecht et al., 1992).

IV. SHIP-GENERATED WAVES

As water flows back past a moving vessel, a variable pressure distribution develops along the hull. The pressure rises at the bow, then falls to below the free stream pressure over most of the side of the hull and rises again at the stern. The water surface profile along the hull responds to this pressure distribution causing a rapid rise and fall of the water surface at the bow and to a lesser extent (owing to flow separation) at the stern. Inertial effects cause the water surface to "overshoot" its equilibrium position and generate sets of waves that propagate out from the bow and stern of the vessel. The hull pressure distribution and resulting height of generated waves depend on the water velocity relative to the hull (i.e., the ship speed and the clearance between the hull and the channel bottom and sides) and the bow and stern geometry. The periods and propagation directions of the generated waves depend only on the ship speed and the water depth.

For a detailed discussion of the process of ship wave generation and the resulting wave characteristics, see references section (Robb, 1952; Sorensen, 1973a). Only an overview of the pattern of ship waves and their characteristics are provided in this chapter.

A. Ship Wave Patterns

Fig. 10.7 shows the pattern of wave crests generated by the bow of a ship moving over deep water. It consists of symmetrical sets of diverging waves that move obliquely out from the sailing line and a single set of transverse waves that move in the same direction as the ship. The transverse and diverging waves meet to form cusps located along lines that are 19 degrees 28 minutes out from the sailing line independent of the hull design and ship speed. The largest wave amplitudes are found at the cusp points. If

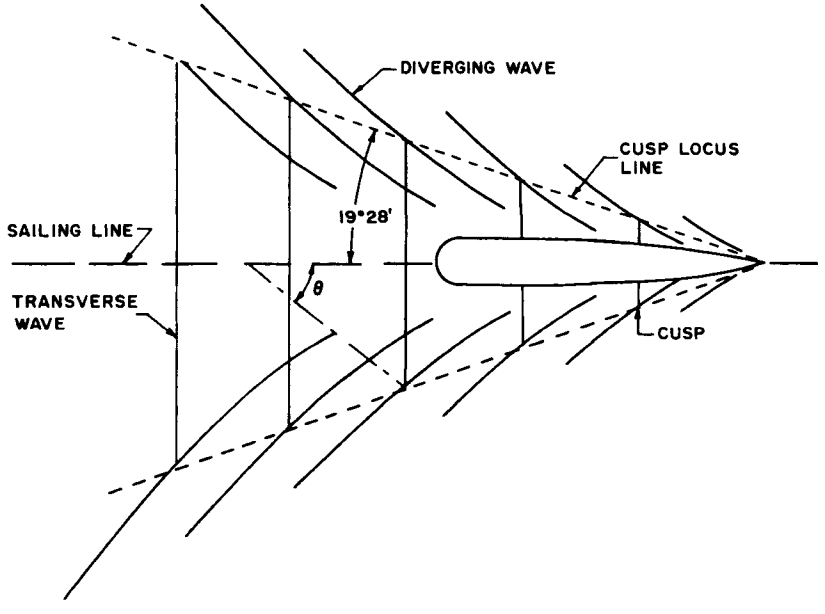


Figure 10.7.—Deepwater Wave Crest Pattern Generated at the Bow of a Moving Vessel.

the ship speed is increased, this pattern retains the same form but expands in size as the individual wave lengths increase. A similar pattern of waves, typically with lower amplitudes, would be generated at the ship's stern.

Since the wave system remains steady with respect to the ship as it moves forward at a constant speed, the waves that form the pattern must have a phase celerity C given by

$$C = V \cos \theta \tag{10.32}$$

where V is the ship speed and θ is the angle between the sailing line and the direction of wave propagation (Fig. 10.7). For diverging waves in deep water, the theoretical value for θ is 35 degrees 16 minutes.

For a given ship speed, at increasing distances from the bow, diffraction causes the wave crest length to increase and, consequently, the wave height to decrease. Havelock (1908) demonstrated analytically that the wave heights at the cusp points should decrease at a rate that is inversely proportional to the cube root of the distance from the bow, while the transverse wave heights along the sailing line decrease at a rate that is inversely proportional to the square root of the distance from the bow. Thus, at greater distances from the ship, the diverging waves become relatively higher than the transverse waves.

Wave amplitudes increase exponentially as the ship speed increases. When ship speed increases to a point where the length of the diverging waves exceeds twice the water depth, the waves "feel" the channel bottom and their characteristics and the wave crest pattern begins to change. This "shallow water" condition occurs for a vessel Froude number, F, in excess of approximately 0.7 where

$$F = V/\sqrt{gd}, \tag{10.33}$$

d is the still water depth, and g is the acceleration of gravity.

In shallow water, as the Froude number increases from 0.7 to 1.0, several changes occur. Transverse wave heights increase at a faster rate than the diverging wave heights, so they become relatively more prominent as the Froude number approaches unity. The cusp locus angle increases from the deep water value of 19 degrees 28 minutes to 90 degrees at a Froude number of unity. At F = 1, with a cusp locus angle of 90 degrees, the diverging and transverse waves have coalesced with their crests oriented perpendicular to the sailing line. Also, most of the energy in the wave system is concentrated in the first wave at the bow. Self-propelled vessels cannot exceed a Froude number of unity and most vessels operate at speeds such that their Froude number is less than 0.9 (Schöfield, 1974).

Very light vessels, moving at increasing speeds, may develop sufficient hydrodynamic lift to cause them to plane. After the onset of planing there is usually no significant increase in generated wave heights for further increases in the vessel speed (Sorensen, 1973b).

B. Ship Wave Characteristics

For the design of bank protection works or marina mooring areas, expected ship wave conditions must be known as a function of vessel draft, speed, and distance from the point of concern. The wave conditions of primary interest are the dominant approach direction, the dominant wave period and a meaningful representative wave height. When designing for wind waves, a representative height commonly used is the significant height. Vessel-generated wave spectra differ from wind wave spectra and have not been studied in sufficient detail to define an appropriate representative height.

The highest ship-generated waves are at the cusp point and have the direction and period of the diverging waves at that point. Based on the theoretical limits of 35 degrees 16 minutes and 0 degrees for deep water and travel at a Froude number of unity along with available empirical data, Weggel and Sorensen (1986) developed the following relationship for the direction of wave propagation θ (in degrees):

$$\theta = 35.27\,(1 - e^{12(F-1)}). \tag{10.34}$$

Thus, for a given vessel speed and water depth the wave direction can be determined from Equation 10.34 and then the diverging wave celerity can be determined from Equation 10.32. Then, the diverging wave period T can be determined from a modified form of the wave dispersion equation

$$C = \sqrt{\frac{gCT}{2\pi} \tanh \frac{2\pi d}{CT}} \tag{10.35}$$

by trial and error. For common water depths and ship speeds, the diverging wave period has typical values between one and three seconds.

More than a dozen laboratory and field investigations of ship-generated waves have been conducted during the past three decades (Sorensen and Weggel, 1984; Weggel and Sorensen, 1986). Typically, the water surface time history was measured for each vessel passage to yield a record similar to that in Fig. 10.8. Usually, the maximum wave height (H_m in Fig. 10.8) from each record was reported along with the vessel speed and type, the water depth, and the distance from the ship's sailing line. An occasional author has presented an average or a root mean square wave height from each record, but these data are quite limited. Sorensen (1967; 1973b) reports on measured values of H_m for a variety of vessel sizes and shapes in open water. For ship speeds between 6 and 14 knots, H_m values range from 0.2 to 0.9 meters at a distance of thirty meters from the sailing line and from 0.1 to 0.8 meters at a distance of 150 meters from the sailing line. These values are representative of H_m values found in most common navigation situations.

At least six procedures have been developed for the prediction of H_m. They are all quasi-empirical and most are restricted to a specific range of vessel and/or channel conditions. Sorensen (1989) references these procedures and briefly discusses them. A comparison calculation of H_m for a selected ship speed, water depth, and distance from the sailing line showed significant deviation of results. The best approach appears to be to use the method(s) that most closely compares to design conditions and temper the result with field measurements discussed above.

V. WAVE-STRUCTURE INTERACTION

Water waves and their kinematics are important in hydrology and hydraulic engineering. Waves interact with boundaries of estuaries, causeways, tidal inlets/entrances, river embankments, coastal

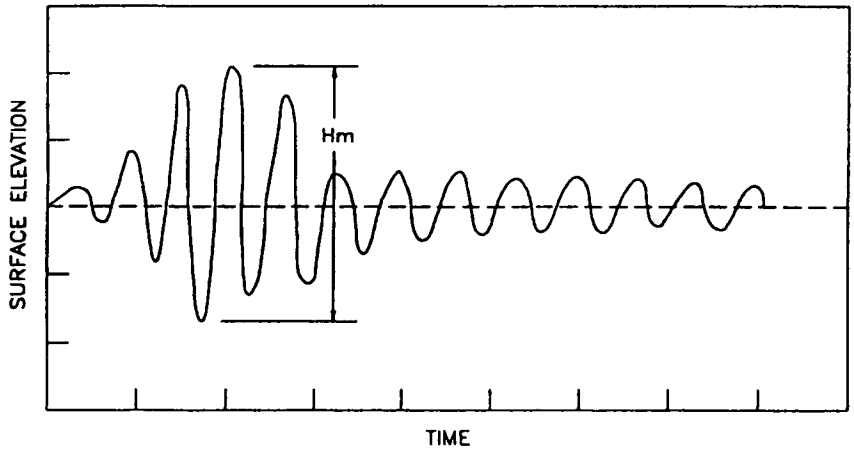


Figure 10.8.—Typical Ship-Generated Water Surface Time History.

beaches, marinas, and small boat harbors. These and other man-made hydraulic structures are built for attenuating wave impact and providing a safer and more navigable environment in the waterways. The ability to quantify the interaction of regular and random waves with surrounding lands and structures is essential in design work. This interaction is often described in terms of parameters such as wave runup or rundown, wave overtopping of structures, wave transmission and reflection coefficients or sometimes in the form of fluid loads acting on structures. The literature includes a large number of publications on each of these topics, and therefore only a brief overview of the subject will be presented.

A. Regular Wave Runup and Rundown

Reflection of waves by structures can cause significant fluctuations in the water level. Similar variations in water level may also occur when waves travel up inlets or entrances to waterways and estuaries, or lakes and reservoirs. Incident waves generated by winds, ships, or sliding lands are modified by structures and land boundaries through wave reflection or transmission and wave breaking. The difference between a wave breaking on a structure and a gentle beach is the steep slope of the former. As waves break on a slope, the wave momentum pushes the wave profile to rush up the slope. The runup, R_u, is the local maximum or peak vertical elevation reached by wave elevation on the face of a structure (Fig. 10.9). This maximum elevation includes both the wave set-up, the superelevation of the mean water level caused by the energy of waves known as the radiation stress, and the swash that represents the fluctuations about that mean. The runup depends both on slopes of the bed and structure, and reflection from these slopes. For wind waves, runup is a function of the incident wave characteristics, including wave height, wave period, wave steepness, and the distribution of wave energy in the wave energy spectrum. The water depth at the toe of a structure and wave breaking also influence the runup. For a fully-reflective and impermeable vertical wall, wave runup is theoretically twice the incident wave height.

For regular waves, an empirical relationship may be derived between runup and wave properties, geometry of a structure, water depth, slope angle, structural surface roughness, porosity, and beach face slope. Wave breaking on slopes reduces the energy of incident waves through reflection and runup. The mechanics of wave breaking and degree of reflected and consumed momentums are controlled by the steepness of waves (H/L) and slope of the beach ($\tan\alpha$). Using the subscript 0 for values of deep water wave height and wavelength, these may be combined into one dimensionless variable called the "surf similarity parameter" or "Iribarren's number" defined comparable to Equation 10.20 as

$$\xi = \tan\alpha/(H_0/L_0)^{1/2}. \tag{10.36}$$

The influences of toe depth and slope are obvious; steep slopes cause greater reflection, resulting in little or no runup. On relatively flat or gentle slopes, the up-rush of waves is retarded by the bottom friction over the distance water travels, limiting the vertical reach of water usually to be less than the length of the sloping surface. On mild slopes, it is generally difficult to define the wave runup precisely due to either the foamy water front after wave breaking, or the presence of a very thin film of water on

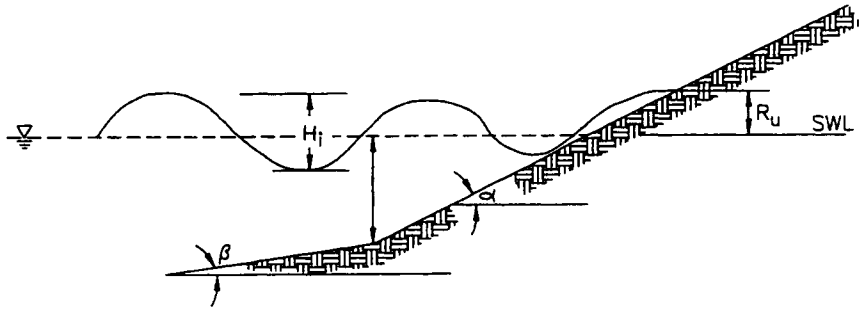


Figure 10.9—Definition Sketch for Wave Runup on a Sloping Structure

the slope. In this case, runup may either be taken as the maximum or some average rush-up water level. This definition gives rise to a large difference between runup measurements from laboratory and field studies. The temporal error in the measured runup at any time depends on how accurate the position of water elevation on a structure may be located. Waves may break before reaching the toe of a structure or while on its sloping surface. Thus wave breaking is the essence of runup and rundown calculations, and will next be briefly discussed.

The height of a wave moving over a sloping beach may increase while its length decreases; waves are deformed by a shoaling process before breaking. Thus, the steepness of waves, defined as the ratio of wave height to wavelength H/L, approaching a beach increases. Waves break when they reach a limiting steepness. In general, wave steepness depends on the relative water depth defined as d/L, and the beach slope, tanα. The theoretical wave breaking limit for constant water depth is (Miche, 1944)

$$(H/L)_{max} = 0.142 \tanh kd. \tag{10.37}$$

This equation may also be used for very mild slopes.

The form of waves at breaking is known as the breaker type, characterized by spilling, plunging, collapsing, and surging types. The surf similarity parameter, based on deepwater wave assumption (i.e., $d/L > 1/2$), can be used to distinguish breaker types (Galvin, 1967; Weggel, 1972; Battjes, 1974a). Wave breaking types as a function of the surf similarity parameter are listed in Table 10.4.

Waves of very low steepness on a slope may not break but instead reflect from the slope to form a standing wave in front of the slope. In general, steep waves on flat beaches or low slopes spill; waves of moderate steepness plunge on steeper slopes, while low steepness waves surge on the steeper slopes. Plunging breakers produce high local impact pressures on sloping surfaces. Winds also influence the breaking of waves; onshore winds tend to trigger breaking in deeper depths whereas offshore winds retard breaking, forcing it to occur in shallower depths.

If waves do not break until reaching a structure, the sloping face of the structure may itself cause waves to break. The slope of a structure increases the breaking wave height and water depth at the point of breaking as (Le Mehaute and Koh, 1967)

$$H_b/H_0 = 0.76(\tan\alpha \cos\alpha_b)^{1/7} (H_0/L_0)^{-1/4} \tag{10.38}$$

where H_b is the breaking height at the toe of structure, H_0 and L_0 are the deepwater unrefracted incident wave height and wavelength, respectively, and α_b is the breaking angle to the shore. The deepwater wavelength may be estimated in units of meter by $L_0 = 1.56\ T^2$. Collins (1970) shows that depth at breaking is roughly $1.3H_b$. For slopes less than 1 on 8, this may be estimated by

$$H_b/d_b = 0.72 + 5.6 \tan\alpha \tag{10.39}$$

Different formulas have subsequently been derived for breaking wave height and water depth (Weggel, 1972; Komar and Gaughan, 1973) from laboratory studies on smooth, plane slopes using linear wave theory (USACERC, 1984). For irregular waves, the breaking root-mean-square (rms) wave height is approximately half the local water depth at the breaking point.

The wave runup elevation on a slope is different for nonbreaking and breaking wave conditions. In

TABLE 10.4. Classifications of Wave Breaking.

Surf Similarity Parameter ξ	Wave Breaking Type
0.5 or less	Spilling
0.5 to 3.3	Plunging
3.3 or larger	Surging and Collapsing

the absence of information about wave breaking, runup may be estimated for both nonbreaking and breaking waves, using the following respective formulas (Hunt, 1959):

$$R_u = 2\, a_c\, A \tag{10.40}$$

$$R_u = (H_0 L_0)^{1/2} \alpha \tag{10.41}$$

in which a_c is the incident wave crest elevation above the still water level, and α is the face slope in radians. The parameter A varying with $K = (kd/\alpha)(L/L_0)$ is the surface amplification factor (Fig. 10.10) that determines if waves break on a slope.

When the toe of a structure with a steep slope is placed in deep water, the nonbreaking wave runup may be approximated by (Miche, 1944)

$$R_u = H_0\, (\pi/2\alpha)^{1/2} \tag{10.42}$$

Wave steepness generally increases the runup elevation on a structure with a sloping face. This increase may be estimated from the standing wave theory (Miche, 1944). The resulting wave runup is then determined from (Takada, 1970; 1977)

$$R_u = H_0\, K_s \left[(\pi/2\alpha)^{1/2} + (R_v/H_t - 1)\right](\cot \alpha_c \tan \alpha)^{2/3} \tag{10.43}$$

$$R_v/H_t = 1 + (\pi H_t/L_t)\coth k_t d_t \left[1 + 0.75\, \mathrm{cosech}^2 k_t d_t - 0.25\, \mathrm{sech}^2 k_t d_t\right] \tag{10.44}$$

where H_0 is the deep water wave height, R_v is the wave runup elevation on a vertical wall ($\alpha = \pi/2$), K_s is the shoaling coefficient (USACERC, 1984), H_t is the wave height at the toe of the slope, and α_c is the limit angle for waves not to break on a slope. Here k_t, L_t, and d_t are the wave number, wavelength, and water depth at the toe location.

Data on the effect of wave steepness on runup is available (USACERC, 1984) for various ratios of d_s/H_0' where $d_s = d_t$ and $H_0' = H_0$. Samples of the theoretical and experimental runup estimates for different values of structural slopes fronted by a 1 on 10 beach slope are summarized in Figs. 10.11, 10.12, and 10.13.

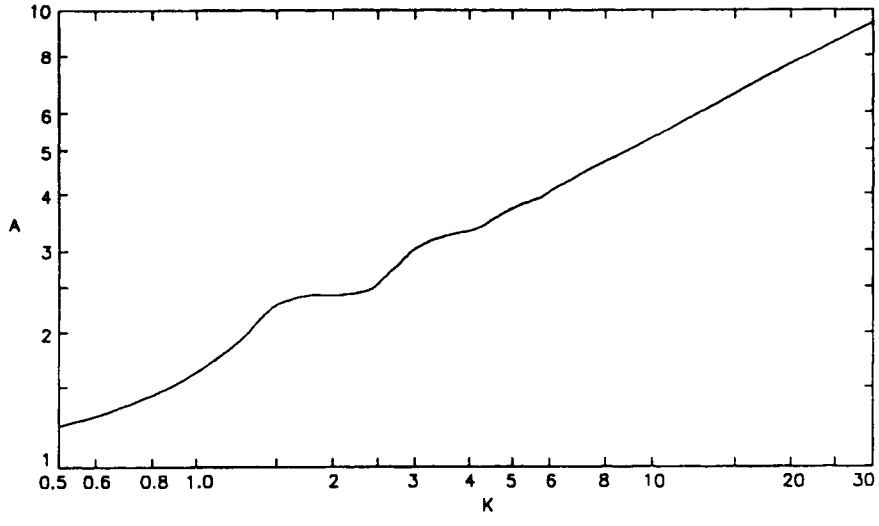


Figure 10.10.—Amplification Factor for Wave Runup Estimates

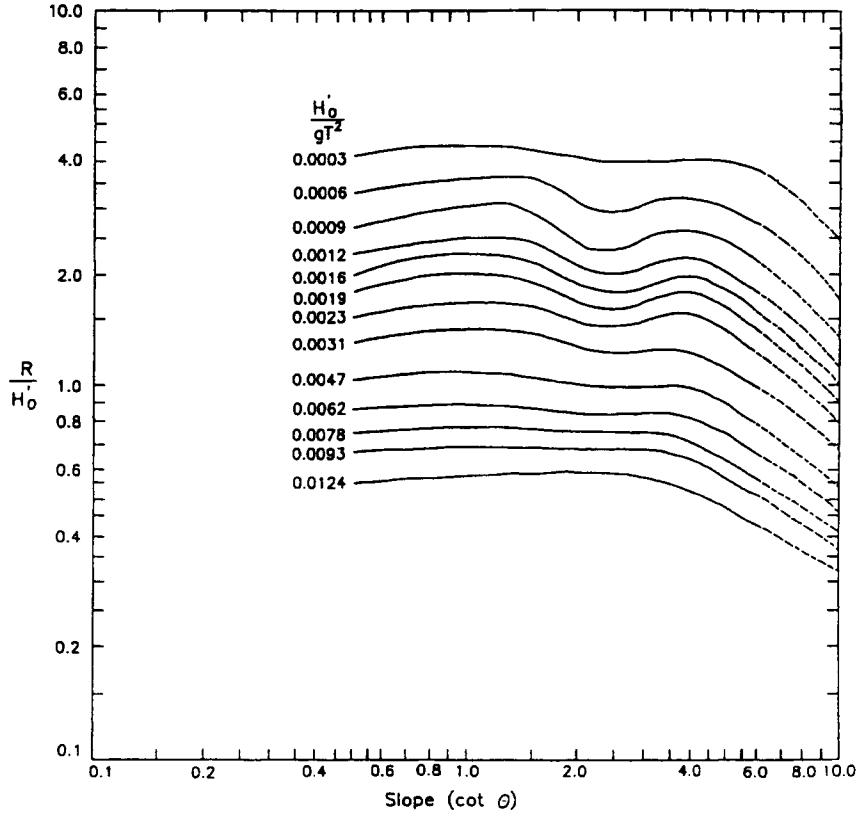


Figure 10.11.—Wave Runup on Smooth, Impermeable Structural Slopes Fronted by a 1:10 Beach Slope for $d_s/H_0 = 0.0$ (USACERC, 1984).

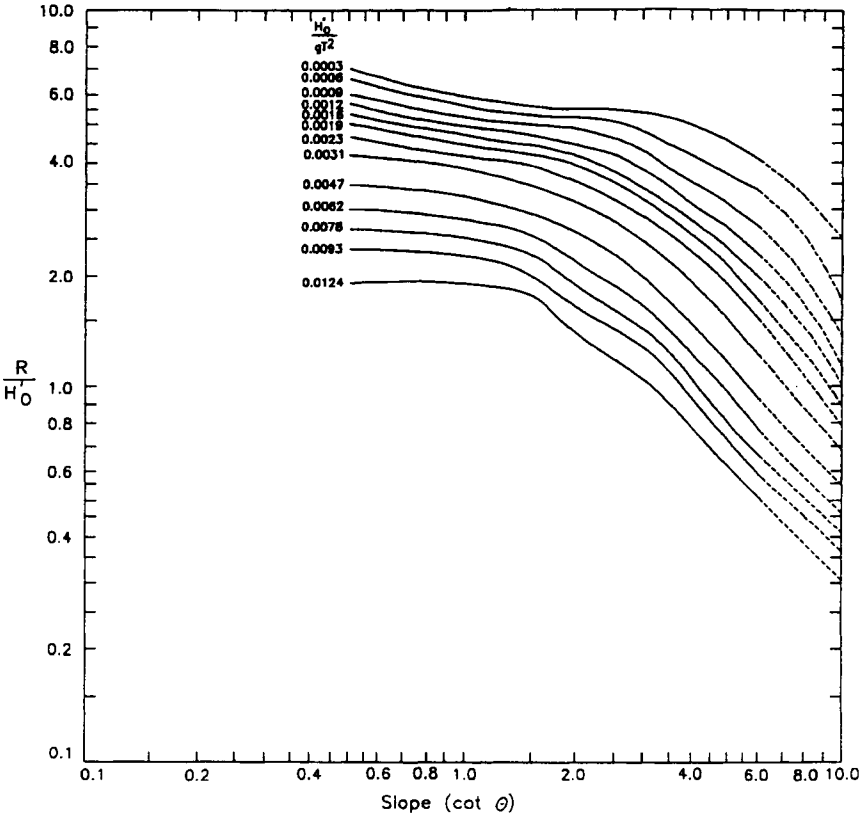


Figure 10.12.—Wave Runup on Smooth, Impermeable Structural Slopes Fronted by a 1:10 Beach Slope for $d_s/H_0 = 0.8$ (USACERC, 1984).

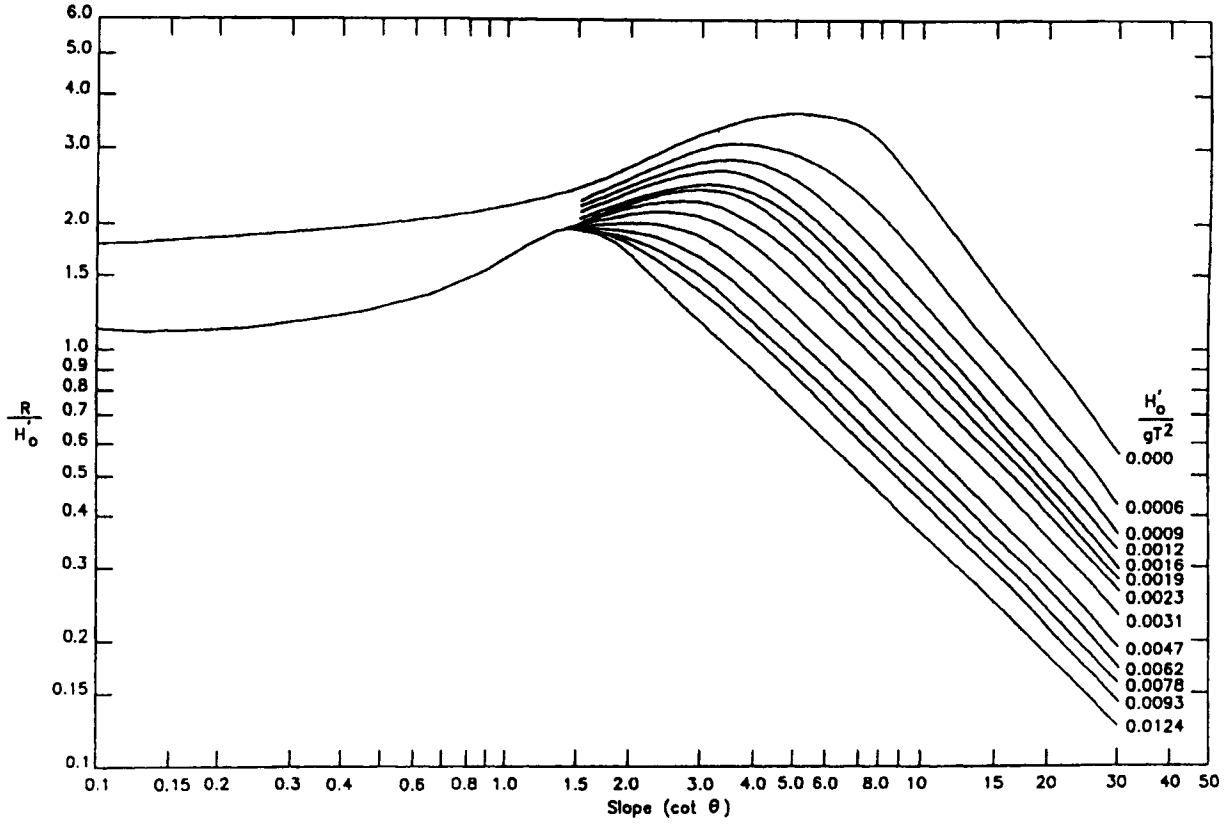


Figure 10.13.—Wave Runup on Smooth, Impermeable Structural Slopes Fronted by a 1:10 Beach Slope for $d_s/H_0 \geq 3.0$ (USACERC, 1984).

When waves break on the slope, the following equation has been derived from experimental results for runup estimates (Hunt, 1959):

$$R_u = 1.01\, H_t \tan \alpha\, (H_t/L_0)^{-1/2} \tag{10.45}$$

To determine whether or not overtopping is a serious problem in the design of mound structures (riprap and rubble mound) requires the prediction of wave runup on these irregular shape structures. Most of these protective structures have rough and highly permeable surfaces which absorb wave energy and reduce wave runup. The fundamental difference between runup on rubble mound structures and riprap revetments is in their structural design. On rubble mound slopes, runup and rundown exhibit characteristics similar to the smooth slope results reduced by a coefficient representing effects of surface roughness and permeability (USACERC, 1984). Comparison of wave runup on smooth slopes with runup on permeable rubble slopes is presented in Fig. 10.14.

A correction factor, r defined as the ratio of runup on a rough permeable slope or other slopes to the runup on a smooth impermeable slope should be used when applying smooth-slope design curves to the roughened and porous structures. This factor is specified in Table 10.5.

The determination of wave rundown for regular waves on sloping surfaces often is based on the assumption that rundown on these surfaces occurs on a wedge profile. The water surface runs down the wedge, varies with time, and attains its maximum velocities before the face slope is dry. Based on experimental data, a formula has been suggested (Roos and Battjes, 1976) for rundown on smooth slopes. This formula does not allow rundown to penetrate below the still water level at low values of ζ.

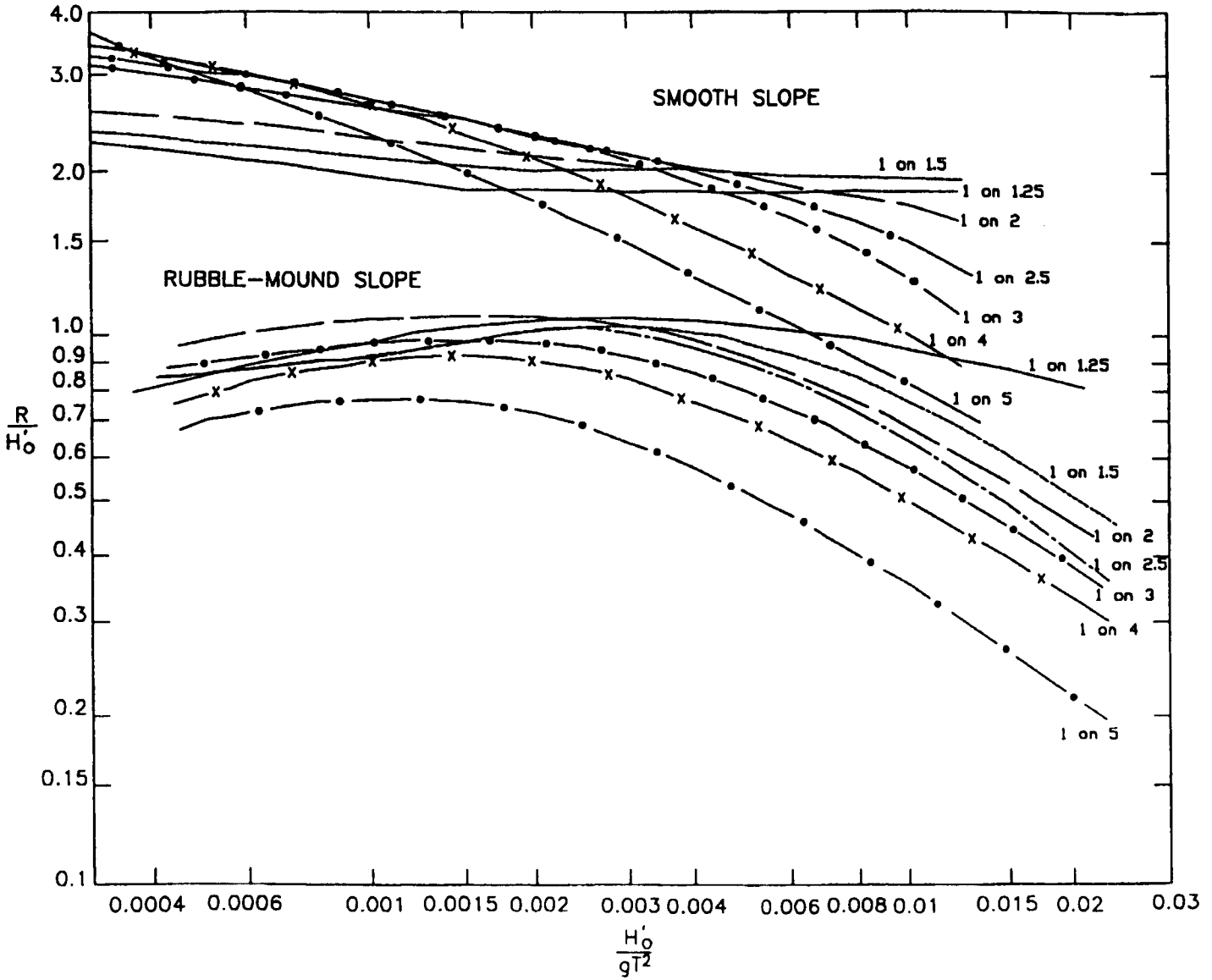


Figure 10.14.—Comparison of Wave Runup on Smooth Slopes with Runup on Rubble Slopes for $d_s/H_0 \geq 3.0$ (USACERC, 1984).

TABLE 10.5. Roughness and Porosity Correction Factor.

Sloping Surface	r
Smooth	1.0
Concrete, Basalt and Gobi Blocks	0.85–0.90
Grass	0.90
Quarrystone (one layer)	0.70–0.80
Quarrystone (rounded)	0.65–0.70
Quarrystone (layers)	0.50–0.60
Concrete Armor Units (50% void ratio)	0.40–0.50

Assuming that rundown is independent of wave height on smooth slopes for low values of ξ, say for $0.3 < \xi < 3.0$, an empirical formula curve-fitting the model data for rundown is

$$R_d/H = (1 - 0.45\xi)\xi \tag{10.46}$$

Physical model results are correctly represented by this equation (Bruun, 1985). Fig. 10.15 presents wave rundown and wave runup information for a graded riprap with slope 1 on 2 over an impermeable base.

B. Irregular Wave Runup and Rundown

The majority of runup and rundown results have been obtained from tests with regular waves. Hydraulic structures in nature are exposed to irregular waves, and therefore design estimates should be based on irregular sea states. In this case, since irregular wave properties vary randomly, it is necessary to treat runup and rundown as stochastic variables which depend on wave characteristics for a given slope, including the wave energy spectrum and its statistical parameters. The preferred approach in practice has been to relate the wave runup and rundown of irregular seas to their counterparts obtained from regular waves. The transfer of results from regular waves to irregular waves uses the so-called "hypothesis of equivalence" (Saville, 1962; Van Oorschot and d'Angremond, 1968; Battjes, 1971; Gunbak, 1977, 1978) that assumes each wave component in a spectrum generates runup/rundown values which would be similar to those by a regular wave of the same height and period.

In hydraulic design studies, a wave spectrum may be approximated by a monochromatic wave of period T_p which is the period associated with peak spectral energy. Consequently, for irregular waves, the significant wave height, $H_{1/3}$ (or H_s) and T_p should be used in the formulas and figures for regular waves. A conservative estimate of the component wave runup resulting from irregular waves is plausible assuming that wave runup may be represented by the Rayleigh distribution. The probabilistic runup estimate then becomes (Van Oorschot and d'Angremond, 1968; Battjes, 1974a, 1974b)

$$R_p/R_s = H_p/H_s = [\ell n(1/p)/2]^{1/2} \tag{10.47}$$

in which R_p is the runup for a "p-percent" probability of exceedance (i.e., the elevation exceeded by p-percent of runups), and R_s is the runup for H_s. As an example, $R_{0.01}$, or the 1% runup, would be 1.52 times greater than R_s (Fig. 10.16). The percent runup means the percent exceedance, and it should not be confused with the average value for a given percent probability.

C. Wave Overtopping

When the crest elevation of a structure is lower than the maximum runup, water will spill onto or over it. Overtopping does not refer to some spray or splash over a structure; it represents overrun by clear water up-rushing the structure. For quantifying the consequences of overtopping, it is customary to define overtopping as the volume of flow per unit time and crest length. This may also be expressed as a damage criterion for sea walls, a required armor units weight criterion for mound structures, or a water

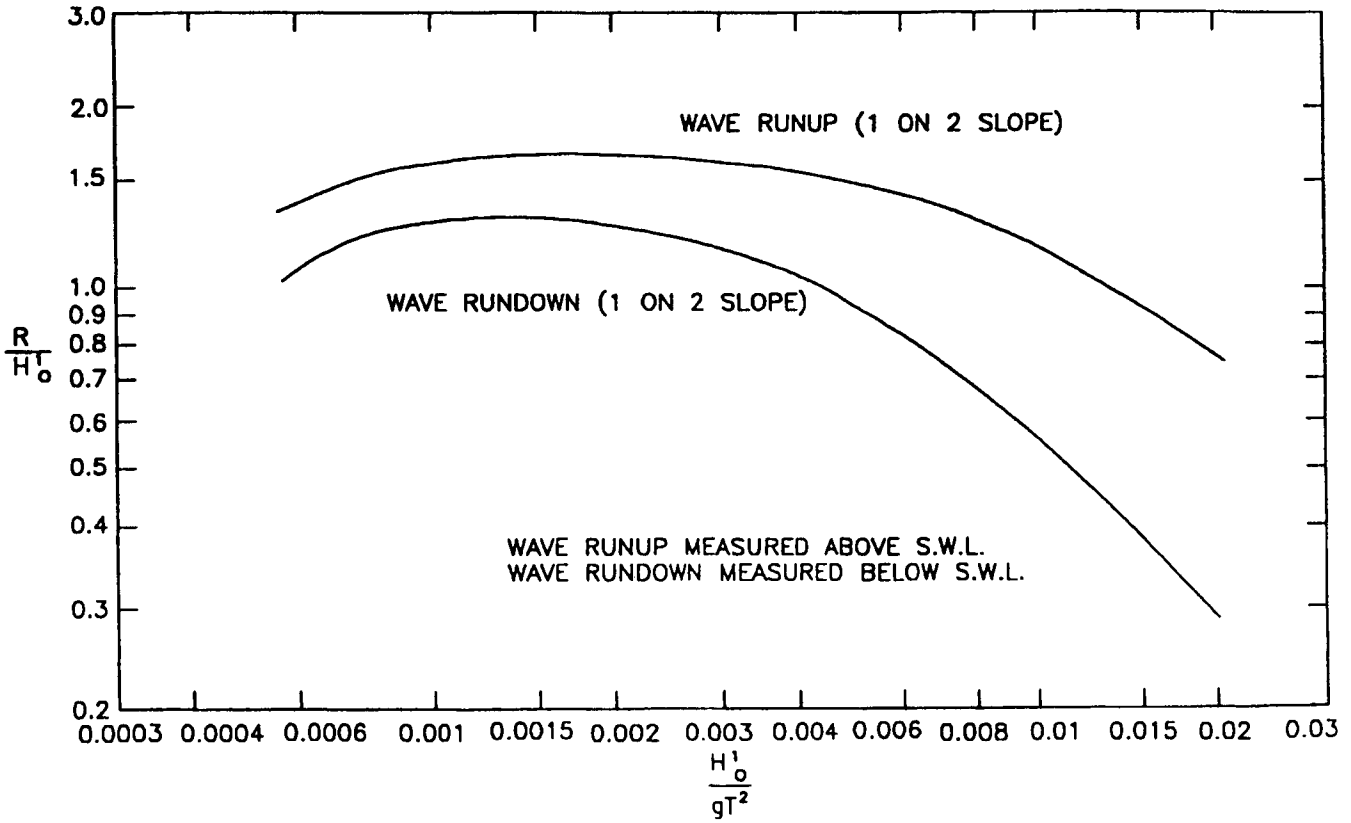


Figure 10.15.—Wave Runup and Rundown on Graded Riprap of Slope 1:2 on an Impermeable Base for $d_s/H_0 \geq 3.0$ (USACERC, 1984).

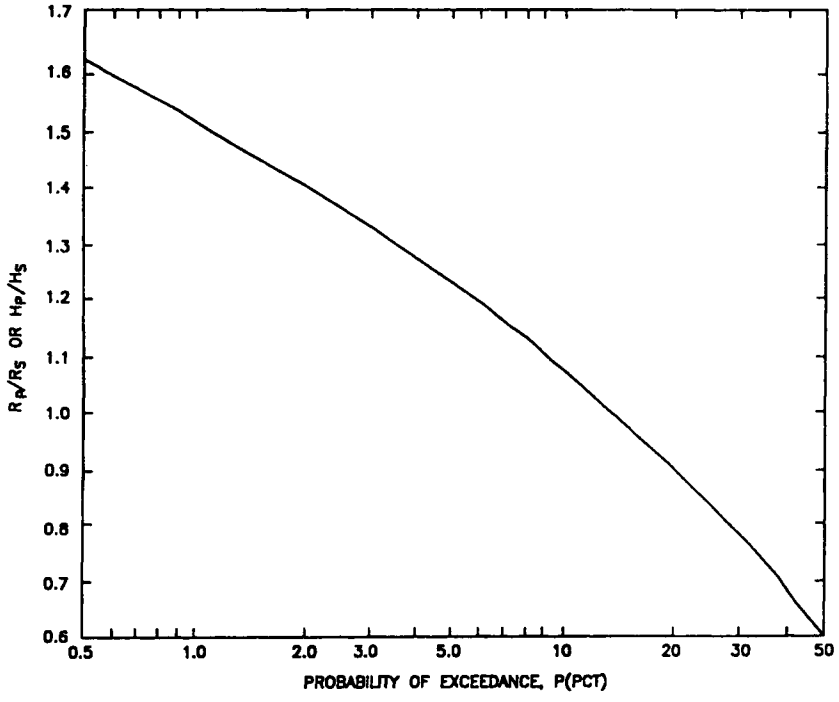


Figure 10.16.—Relative Runup R_p/R_s or Relative Wave Height H_p/H_s as a Function of the Probability of Exceedance p (Battjes, 1974).

removal criterion for sizing drainage systems. In general, it is prohibitively expensive to build hydraulic structures to preclude overtopping from the largest waves, and consequently, a certain quantity of overtopping flow may have to be allowed in the planning and design. The quantity of overtopping can be established beforehand from the extent of a flood area and by the required capacity of pumping facilities for dewatering the downstream side of the overtopped structure.

The prediction of overtopping for regular and irregular waves is described in detail in SPM (USAC-ERC, 1984). The factors that have the greatest influence on overtopping include the face slope and water depth at the toe of a structure and the wind speed. In general, the overtopping rate increases with increasing face slope of a structure, attaining the maximum rate when slope is roughly 1:2 to 1:3. The toe depth determines whether or not the structure is positioned in the standing or the breaking wave zone. The wind supplies more energy into waves, and thus overtopping increases proportionally to the wind speed. Recent research studies (Takada, 1970; 1977; Sato and Goda, 1981) have modified the earlier results (Saville, 1956; 1958; 1960; Savage, 1958) to significantly improve overtopping predictions. The following formulas may be used for estimating overtopping of regular and irregular waves, respectively:

$$Q_m = (gQ_0H_0'^3)^{1/2} \exp[-u] \tag{10.48}$$

where for $0 \leq (h-d_s)/R_u \leq 1.0$, and

$$u = (0.217/\alpha) \tanh^{-1}(h-d_s)/R_u \tag{10.49}$$

$$Q_i = [gQ_0(H_0'^3)_s]^{1/2} \exp[-uR_u/(R_sK_i)] \tag{10.50}$$

$$K_i = R_p/R_s = (-\ln p/2)^{1/2} \tag{10.51}$$

$$Q_r = 1/199 \, \Sigma \, Q_i \tag{10.52}$$

Q_m is the overtopping rate due to regular waves or volume per unit time per unit structure length; Q_i is the overtopping rate due to irregular waves or volume per unit time per unit structure length associated with a particular value of K_i and $(H'_0)_s$, where $(H'_0)_s$ is the equivalent deep water significant wave height; index "i" refers to ranges of probabilities desired to be generated for runup with a Rayleigh distribution; index "p" refers to particular probability of exceedance; R_p is the runup associated with a particular exceedance limit; R_s is the runup corresponding to a deep water significant wave height and period, predicted by conventional regular wave studies; $Q_{1/2}$ is the average overtopping rate for runup with a Rayleigh distribution having a probability of exceedance of p = 0.005 or 0.5%; g is gravitational acceleration; H_0' is the equivalent deep water wave height; h is the height of the structure crest above the bottom; d_s is water depth at the structure toe; R_u is runup on the structure if it were high enough to prevent overtopping; $h - d_s/R_u$ is relative freeboard; and α and Q_0 are empirical coefficients (see USACERC, 1984 for values of these parameters).

Note that for a 0.5% exceedance (p = 0.005), the corresponding overtopping rate, $Q_{1/2}$, is for a runup of 63% greater than R_s (K_i = 1.63). Therefore, it is customary to refer to $Q_{1/2}$ as the *peak overtopping rate.* Fig. 10.17 shows a family of curves for ratios of Q_r/Q_m and $Q_{1/2}/Q_m$ for values of α ranging from 0.04

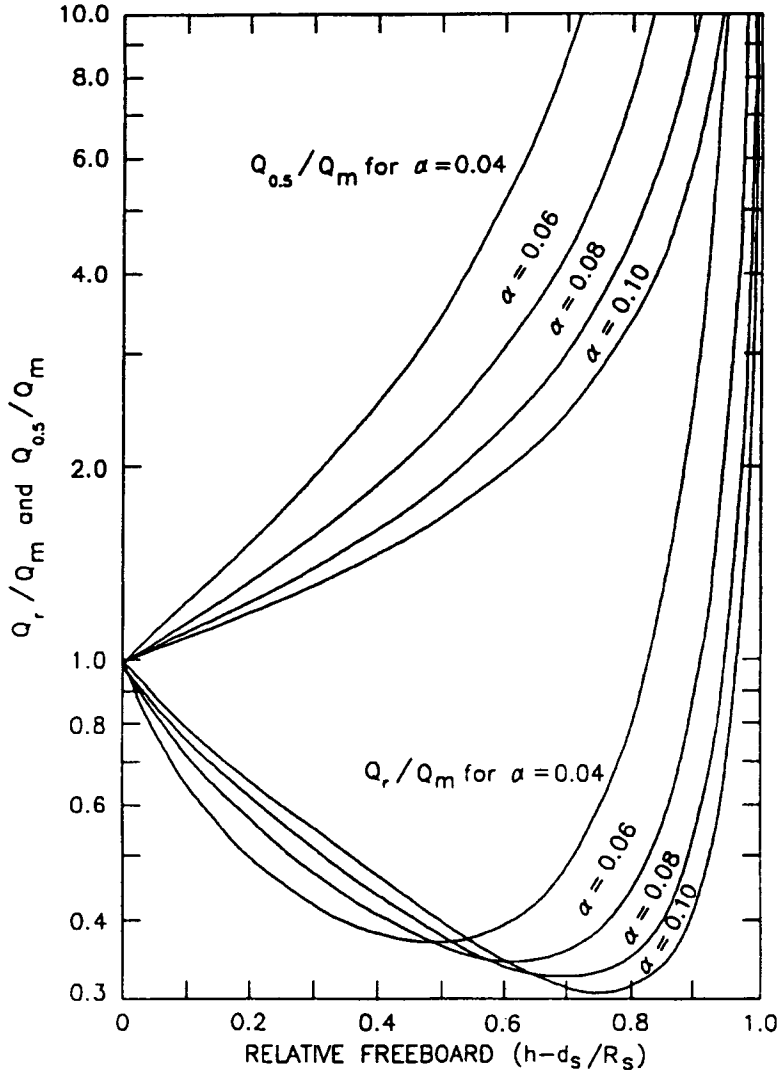


Figure 10.17.—Ratios of the Wave Overtopping Rates (see Eqs. 10.41 through 10.45) as a Function of the Relative Freeboard for Different Values of α *(USACERC, 1984).*

to 0.1. These are typical values of α in applications, and clearly this parameter can strongly influence overtopping estimates. A different method may be used for prediction of the overtopping rate due to irregular waves using the discharge formula for flow over a weir.

The effect of winds on overtopping can be important, though little information is available to determine this effect. A correction is necessary for including the effect of wind in overtopping estimates. This correction may be estimated by (Takada, 1977)

$$Q(v) = \exp(\lambda v/c_0)\, Q(0) \tag{10.53}$$

where Q(0) is the overtopping rate in a windless condition and v and c_0, respectively, are the wind velocity and deepwater wave celerity. The value of λ for nonbreaking waves is 0.112, and 0.088 in the zone between the breaking point and the toe of a structure. An approximate value of λ in the shallower waters where depth is less than the toe depth of a structure is 0.231.

The prediction of overtopping on gravel islands may be based on data and semi-empirical methods applicable to the design of straight ripraps and dikes. This approach assumes that runup height around gravel islands would be the same as the runup height on an equivalent two-dimensional structure, and therefore the variation of overtopping rates around the periphery of gravel islands is neglected.

D. Wave Transmission

The determination of wave transmission past hydraulic structures may be required for some projects. Important parameters for wave transmission estimates include the incident wave characteristics (height H_i and period T), water depth d, crest elevation z_c, and crest width of a structure. The angle, roughness and permeability of the slope may also be important for gentle slopes when the crest height of a structure is greater than 10 m. For a fully- or partially-submerged structure, the controlling parameter is the relative submergence of the crest height, z_c/d. The effect of this parameter on wave transmission coefficient varies with the wave steepness and a large transmission coefficient is usually associated with longer waves. In addition, when the crest of a structure is relatively low compared to the wave height, the resulting overtopping may create quite large waves on the lee side of the structure.

For short waves past a structure whose crest is near the still water level, a more appropriate parameter for quantifying wave transmission is the ratio z_c/H_i, where H_i is the incident wave height. In order to accurately predict wave transmission coefficients due to short waves, it may be necessary to solve three-dimensional hydrodynamic diffraction/radiation equations of the flow field around structures. In the preliminary engineering calculations, the wave transmission coefficient, T_r, defined as the ratio of the transmitted wave height to the incident wave height ($T_r = H_t/H_i$), may be approximated by Tr $\approx$ 0 for $z_c/d \geq 0.75$. For $z_c/H_i \approx 0$, it may be assumed as $T_r \leq 0.5$, while $T_r \geq 0.75$ if $z_c/H_i < -0.5$. Due to lack of reliable data, it is difficult to give specific design guidance for wave transmission around rubble mound structures. This is mainly due to the presence of high porosity and mixed-size rocks. For these structures, wave transmission may either be numerically estimated using finite element or boundary integral techniques (Garrison, 1978; Mei, 1991) or determined experimentally by physical model studies.

E. Wave Forces on Structures

The knowledge and prediction of wave forces is useful for analysis and maintenance of various types of hydraulic structures, including seawalls, bulkheads, and revetments that protect lands from waves, storm surges, and currents. These structures have critical functional roles. For example, groins provide protection against land erosion resulting from waves and currents. Jetties prevent shoaling within navigational channels and control both depth and direction of the flow in river mouths, estuaries, and ports. Breakwaters offer protection to shores, develop beaches, reduce wave energy, and provide safe berthing in marinas, ports, and harbors. The calculation of wave forces on these structures is not a trivial task, and it may be performed by three different theoretical methods.

The first method is the Morison formula, (Demirbilek, 1987) (Dean and Dalrymple, 1991) suitable for estimating the fluid loading on slender pile type structures where the net force on a pile is represented by sum of the inertia and drag components using two experimentally determined semi-empirical coefficients. The method is applicable to the slender structures that cause least disturbance to waves since their size is small compared to the wavelength. The drag loads resulting from vortical flows are important in the design of piles.

For a small and non-slender structure, the drag force is usually negligible compared to the inertia (or mass) force. In this case, a second method called the Froude–Krylov (F–K) hypothesis (Garrison, 1978; Mei, 1991) is applicable for determination of wave forces. The F–K technique is a pressure–area method accounting concurrently both for the effects of added mass and diffraction in the computation of wave loads. Experimentally determined force coefficients are also required for the F–K method, but for a thin, elongated structure, the strip theory may be used to obtain this coefficient. For other situations, a third method is necessary for determining wave forces on hydraulic structures: the diffraction theory. The diffraction theory (Garrison, 1978; Mei, 1991) is a numerical technique that uses boundary integral or finite element or a combination of both methods (hybrid). The third method is more accurate than the first two methods, but it requires large and powerful computers in its numerical implementation. Thus, the first two methods continue to be used in practice. The primary difficulty in applying the first two methods is the selection of appropriate force transfer coefficients for typical hydraulic structures of complex geometry and configuration. These coefficients are available only for simple structures, for example, spheres and cylindrical bodies, and consequently, all three methods are needed for engineering design estimates.

The range of applicability of the three wave force methods for a pile structure is illustrated in Fig. 10.18 where dominant regions of the drag and inertia forces are defined in terms of the Keulegan–Carpenter number ($KC = u_0T/D$) and diffraction parameter ($KA = \pi D/L$). For this purpose, deepwater conditions have been assumed. Here D is a characteristic length of the structure (pile diameter), T is the wave period, u_0 is the maximum horizontal water particle velocity, and L is wavelength. Regions I–III are inertia dominated with region II representing the diffraction theory, while the F–K theory may be used in all three regions if force coefficients are known. Morison's formula should be used in regions IV–VI. Key features of each method will next be described for completeness.

F. Morison Equation

Using experimentally obtained drag and inertia force coefficients and theoretical values of the water particle kinematics, Morison et al. (1950) developed the following equation to calculate wave forces on slender piles:

$$F = F_i + F_d = K_m\, du/dt + K_d\, |u|\, u \tag{10.54}$$

where $K_m = \rho C_m (\pi/4D^2)$, $K_d = 1/2\rho C_d D$, F_i and F_d are the inertia and drag components of the total fluid force per unit length of a pile, ρ is the mass density of water, and u and du/dt are horizontal water particle velocity and acceleration, respectively. Parameters C_m and C_d are experimentally obtained inertia and drag coefficients, and D is the pile diameter. The wave kinematics, u and du/dt, may be approximated by the linear wave theory (Airy theory) as

$$u = (\pi H/T)\, [\cosh k(z + d)/\sinh kd] \cos(kx - \omega t) \tag{10.55}$$

$$du/dt = (2\pi^2 H/T^2)\, [\cosh k(z + d)/\sinh kd] \sin(kx - \omega t) \tag{10.56}$$

where H and T are wave height and period, $\omega = 2\pi/T$, d is water depth, k is the wave number $= 2\pi/L$, and L is the wavelength. Deepwater condition corresponds to $d/L \geq 1/2$ where wavelength is given by $L = L_0$

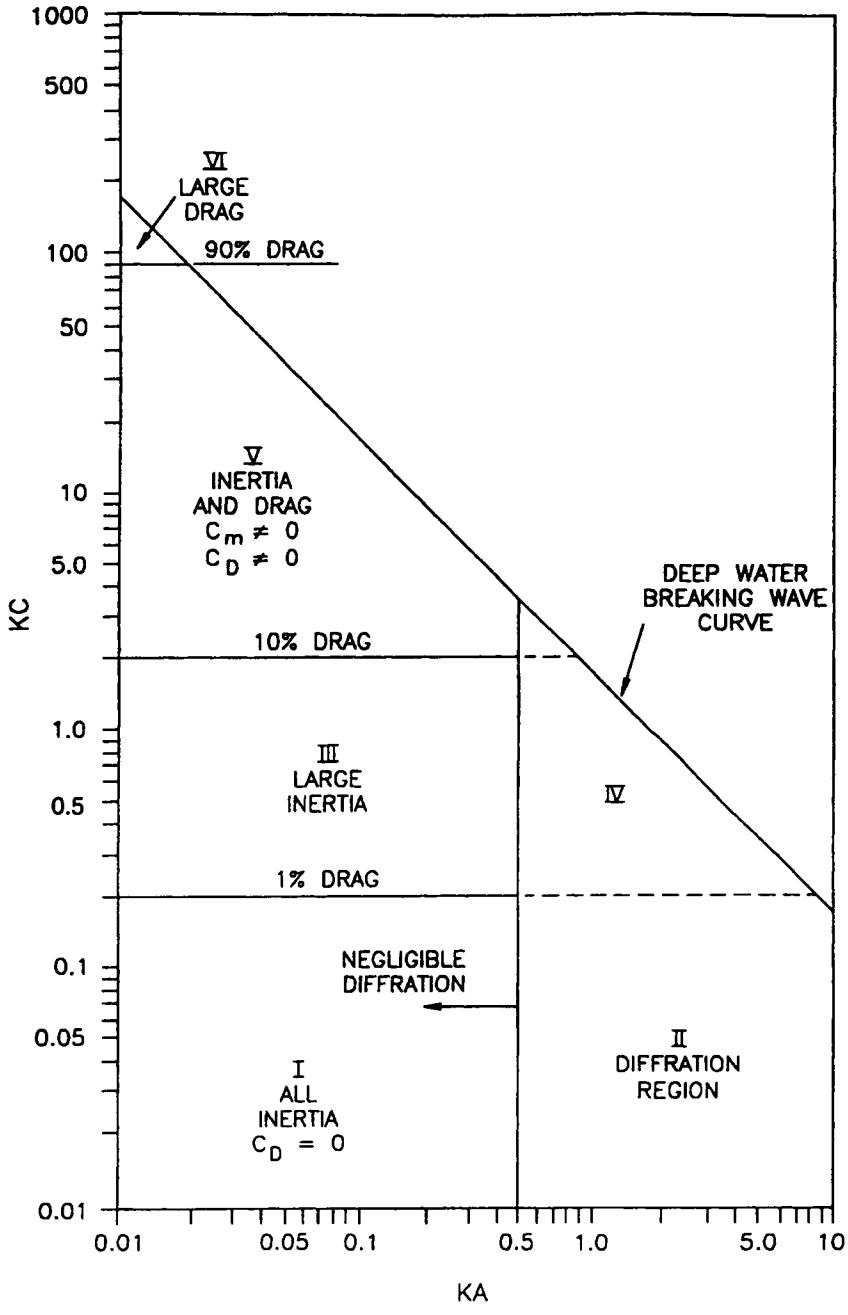


Figure 10.18.—Regions of Application of the Wave Force Formulas (Adapted from Garrison 1978).

$= (gT^2/2\pi)$. For intermediate water depth (i.e., $1/2 > d/L > 1/20$), the wavelength is found from $L = L_0 \{\tanh 2\pi d/L_0\}^{1/2}$, while $L = T\,(gd)^{1/2}$ in shallow water depth if $d/L \leq 1/20$. The two hydrodynamic force coefficients, C_m and C_d may either be determined from laboratory or field measurements made with smooth and rough cylinders. There can be a large difference between these coefficients from similar tests performed at different facilities (Demirbilek, 1987; Chakrabarti, 1987). Values of these coefficients for piles may range from 1.5 to 2.5 for C_m and 0.6 to 1.3 for C_d, respectively. Extreme care is needed when selecting these coefficients for other type structures. Note that Morison's equation can also be used for estimation of wave forces on inclined piles, provided normal velocity and acceleration components are used.

Chapter 11

HYDROLOGIC STUDY FORMULATION AND ASSESSMENT

I. INTRODUCTION

Hydrologic investigations are usually complex, requiring the collection, management, and analysis of large amounts of information. The reliability of study results is always a major concern. Time and financial resources are often limited, but study objectives must be achieved within these constraints. This chapter focuses on four practical considerations in conducting hydrologic investigations: a) formulating the study, b) managing data, c) calibrating and verifying hydrologic models, and d) assessing the accuracy and reliability of study results.

The first topic is concerned with establishing the direction, scope, and resource requirements for a hydrologic study. Clear definition of these parameters for a study will help to ensure that objectives are properly formulated and met and that resources are utilized in an optimal manner.

The second topic deals with the best ways to organize and utilize data. Efficient data management is important because the acquisition and processing of hydrologic data often consume a substantial portion of available study resources. The nature and role of Geographical Information Systems (GIS) are a major focus of this section.

The third topic deals with the problem of identifying and validating values for model parameters. Parameter estimation is a requirement for many types of analysis. For example, aquifer transmissivities are needed for a well yield study, ground water flow rates must be estimated for a contamination study, and soil types and other parameters must be identified to generate surface runoff for a flood analysis.

Finally, this chapter addresses how to evaluate the accuracy and reliability of study results. Concepts and information provided in this chapter are applicable across the range of technical areas covered elsewhere in this handbook.

II. STUDY FORMULATION

Several key elements should be addressed when formulating a hydrologic study, including: a) study purpose and scope, b) level of detail, and c) methods and tools. Each of these elements is discussed in the following sections. Fig. 11.1 is a schematic diagram showing how these elements fit into the formulation process, which culminates in the preparation of a technical study work plan, and leads into performance of a successful hydrologic study.

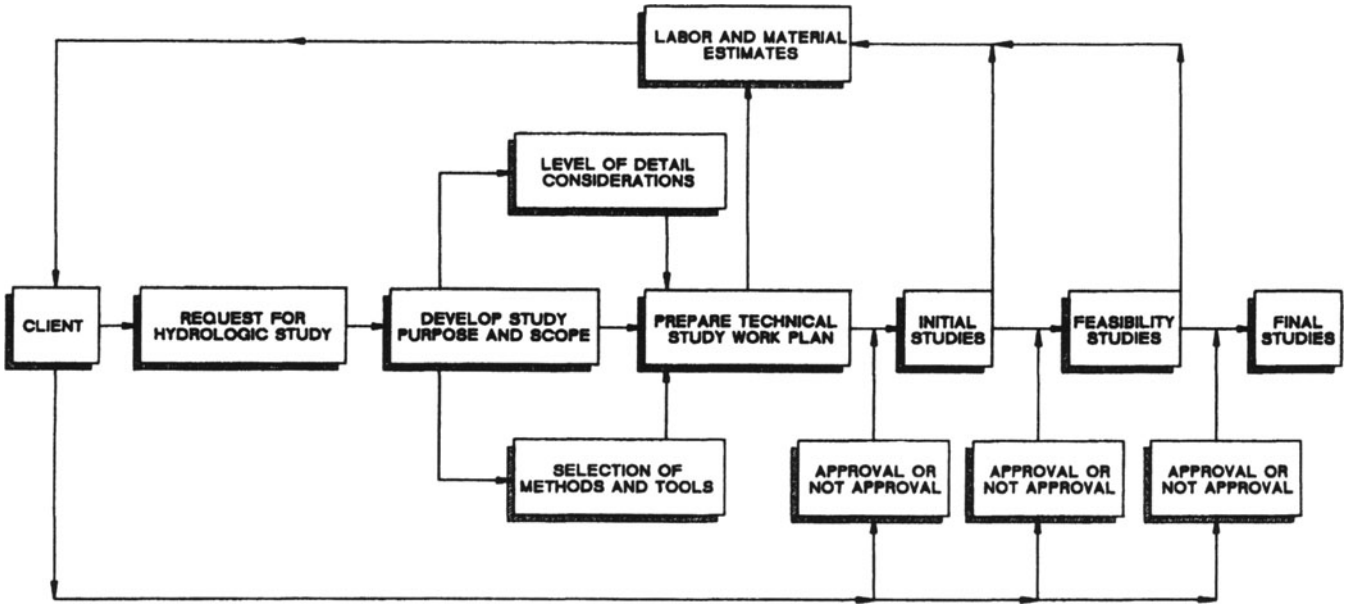


Figure 11.1.—Hydrologic Study Formulation—Process and Stages.

A. Study Purpose and Scope

The typical purpose of a hydrologic study is to provide the technical information that enables decisions to be made on reducing or eliminating existing or potential water-related problems. Examples of problems are:

1) Recurring flood damages in urban and rural areas.
2) Urbanization increases runoff volumes and rates to downstream development.
3) Population growth increases needs for water supply and/or hydroelectric power.
4) Pumping ground water during drought periods depresses ground water levels.
5) Increases in watershed erosion and accumulated sediments result in a loss of reservoir storage or channel carrying capacity.
6) Existing or projected water supply demands exceed reservoir carryover storage.

Hydrologic studies are also conducted to define potential flood hazard areas for regulatory and/or planning purposes.

Generally, the purpose of the hydrologic study will be established prior to the involvement of the hydrologist; however, a hydrologist should be involved in the development of the scope of work, as it must be designed to meet the purpose of the study. The scope of work defines activities required to satisfy the purpose of the study within constraints imposed by schedule, funding, personnel, and/or technical considerations. The scope forms a key part of the Technical Study Work Plan. A tentative scope may be provided by a client, for example, in a request for proposal. Part of the hydrologist's work may involve refining the client's scope to develop a complete scope of work.

There will likely be several specific constraints that must be considered during the development of a scope of work. Examples of constraints are:

1) Time—Planning, design, and construction schedules require hydrology studies to be completed within a specific time period. Institutional concerns and the need for environmental documentation can also affect the time available.
2) Funding—Rarely is a hydrologic study completed without working within funding constraints. Typically, hydrologic studies are funded as part of overall planning studies. These planning funds are generally predetermined based on work previously completed, availability of mapping and data, funds required for similar type and size projects, and estimated construction costs for proposed alternatives to satisfy the study purpose.
3) Personnel—The availability of sufficient qualified personnel who can complete the work within the time schedule may impact the scope of work. If required personnel are not available, it may be necessary to subcontract to other agencies or consultants. In this case, there must be adequate time and funds to advertise, negotiate, and administer outside work activities. It may be necessary to reduce the scope of work to maintain the schedule or negotiate for more time within the original planning schedule.
4) Technical aspects—The methodologies to be used and level of detail of the hydrologic study are very important technical aspects. Both must be appropriate for the time, funding and personnel available, while providing sufficient information for rational decisions. These aspects are discussed in more detail in subsequent sections.

After considering the original purpose of the study and potential constraints, the hydrologist develops a scope that presents the tasks needed to obtain data and apply the selected methodologies at the appropriate level of detail. The scope should include data needs and sources, analyses to be conducted, interactions among different disciplines, documentation to be prepared, and client-coordination points.

B. Level of Detail

Probably the most significant action in formulating a hydrologic study is establishing the level of detail required to satisfy the purpose of each phase of the study. Because the level of detail directly affects the schedule, budget and results of a study, it is vital that an experienced hydrologist be involved at this stage.

The ultimate application of a hydrologic study is a key factor in establishing the appropriate level of detail. Therefore, it is prudent to consider all of the anticipated studies and in some cases the required operation and maintenance activities during the study formulation process. An experienced hydrologist should determine if the hydrologic analyses are needed for a reconnaissance (initial) type of study, for a more detailed feasibility study, or for final design of a project. Obviously, the level of detail required for a reconnaissance type of study is not expected to be as extensive as for a feasibility study or final design.

Typically, during reconnaissance-type studies, the data are inadequate or the funds are insufficient to determine specific design values (e.g., for peak discharges, sediment yields, or runoff volumes). Since the purpose of the reconnaissance-type study is normally to determine whether the project is close to being economically feasible and worth further studies and refinement, solutions with upper- and lower-bound estimates of key parameters may provide a sufficient level of detail. An example is sizing drainage facilities based on estimated peak discharges from similar basins in the region (see Fig. 11.2). A relatively simple analysis (that could be refined during later studies) would be to use upper- and lower-bound peak discharge estimates corresponding to particular drainage areas. A range of sizes for project components would then be determined. If the full range of sizes appears feasible or would not be feasible, then additional refinement is not required because the feasibility question has been answered. Even if the upper- and lower-bound analysis is not definitive, it still provides a better understanding of the sensitivity of a key parameter (peak discharge in this example) to sizing of project facilities.

Even though specific design values would not be generated in a reconnaissance-type study, the level of detail is important. Results from a reconnaissance study are critical, as they may indicate either that a project should be eliminated or that studies should be expanded. Therefore, while expenditures are relatively small for reconnaissance-type studies, their ultimate value can be significant. For example, eliminating a potentially feasible flood control project could have significant and continuing impacts on a community in need of flood protection. Conversely, large expenditures to continue studies on a project that should have been dropped reduces available funds that could have been used on more cost-effective projects.

When selecting the level of detail for each level of study, the use of data and methodology that can be expanded upon or improved as the study progresses should be considered. For example, it may be

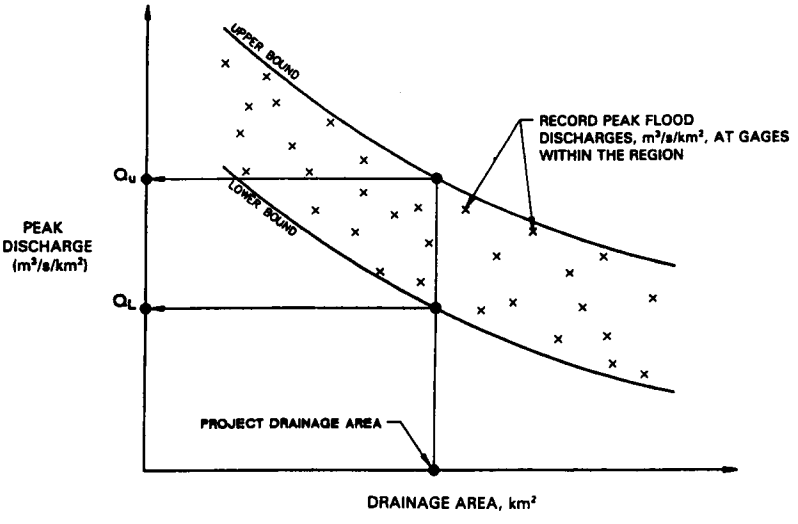


Figure 11.2.—Envelope Curves—Peak Discharge vs. Drainage Area.

reasonable to include rather detailed items such as aerial photography, topographic mapping, field surveys, field samples, and gage readings for a reconnaissance-type study that will later be expanded, because such items can be used at all levels of a study.

The selection of appropriate levels of detail for the various study phases not only facilitates development of the Technical Study Work Plan, but also helps to ensure that only those analyses essential to satisfy the purpose of the study are performed.

C. Selection of Methods and Tools

The selection of methods and tools is particularly important in developing a Technical Study Work Plan. Over the last two decades, the development of computer programs to assist in performing hydrologic analyses has been dramatic. The routine use of desk top computers has brought this technology to essentially all practicing hydrologists, where in past decades only those with access to major computer facilities could utilize available computer programs. The proliferation of computer software provides more alternatives to assist in performing the required studies; however, the wide variety of software makes the selection of the appropriate technology a more difficult task.

Each study may have a number of project-specific items to consider before selecting the methods and tools that best fit the particular study. The following is a list of items that may serve as a starting point for the selection process:

1) project goals and objectives;
2) scope and magnitude of the project;
3) availability of data;
4) difficulty of obtaining essential field data;
5) accuracy requirements;
6) usability of methodology during subsequent studies;
7) staff technical capabilities and training requirements;
8) time constraints; and
9) applicability of computer programs to a specific study (including software documentation and support, and compatibility with other study software packages).

Once the methodology is selected, its application must be clearly defined in the scope of work and incorporated into the Technical Study Work Plan.

D. Preparation of a Technical Study Work Plan

The Technical Study Work Plan is a key document for guiding a hydrologic study. It incorporates the scope of work, which, as discussed previously, defines major work activities and special studies. The work plan should also include a critical path schedule, estimates of required labor and materials, and at least a preliminary content outline of the report(s). Deliverables and milestones should be identified. Key personnel should be shown in a project organization chart, which includes client and other important agency contacts, sub-consultants, and in-house management and technical support staff.

Developing a preliminary outline of the anticipated report(s) can be useful in making sure that all of the major activities have been included in the work plan. Relatively simple flow diagrams and/or bar charts are also useful in preparing and presenting a work plan, as they show the interrelationship of activities and schedule and also serve as a basis for estimating staffing requirements.

Activities that would assist the hydrologist during the preparation of a Technical Study Work Plan include:

1) reviewing available data, mapping, and reports;
2) making site visits, if appropriate;

3) reviewing available methods and tools;
4) coordinating with interested private groups and governmental agencies;
5) evaluating schedule constraints;
6) considering staff constraints;
7) considering funding constraints; and
8) meeting with interdisciplinary team members.

Fig. 11.3 is a schematic diagram that links the activities used to prepare a Technical Study Work Plan to subsequent studies. The intent is to show that the work plan should consider reconnaissance through final studies when developing the total project work plan. Note that the original work plan for studies following the reconnaissance study should be reviewed and refined as more information and reconnaissance study results become available.

The Technical Study Work Plan serves as a guide to the people involved in the performance of the work. With the work plan, the manager responsible for the study has a "blueprint" to work from that is extremely useful in keeping the study focused on critical-path items and on the appropriate level of detail. The work plan is also helpful to the staff working on specific elements of the study by indicating the expected level of effort required, the projected schedule, and where each element fits into the overall study effort.

III. DATA MANAGEMENT

Hydrologic analyses tend to be data intensive. The physical systems being analyzed are often large and complex, so that substantial quantities of data are required for their representation. Also, the analyses themselves are complex, with a variety of interdependent computational elements. The need to transfer information between computational elements can be significant. Because the acquisition, processing, and management of data and information can require a substantial portion of the resources allotted for a study, performance of these tasks in an efficient and reliable manner is necessary.

Hydrologic analyses are performed in both the time and space domains. Data variables routinely treated as time series include precipitation and other meteorologic variables, water discharge of various types (e.g., streamflow, ground water flow, etc.), water stage (or elevation) and storage, water quality parameters, and sediment loads. Spatial variables include topographic and geologic characteristics of basins, soil type, land use, vegetation, and hydraulic characteristics of rivers and flood plains. The following sections deal first with data management considerations in general and then with geographic information systems in particular. The latter are especially powerful for storing and manipulating spatial data.

A. Data Management Concepts

Fig. 11.4 illustrates aspects of data management for a water resource study. Elements of the system include 1) a data-loading module for entering data from various sources into a database, 2) applications software that can retrieve data from, and store results in, the database, and 3) utility programs that perform functions such as data editing, display of data in graphical and tabular form, and mathematical transformation and statistical analyses of data. With such a system, basic data such as precipitation and streamflow can be loaded into data storage and reviewed and perhaps edited with utility programs. Interdependent applications programs can be used to perform the analyses, using the database to pass information between computational elements. Utility programs can be used to view and summarize results in various forms, including report-quality tables and graphs.

There are a number of types of data management systems, each having properties designed to meet specific needs. Because hydrologic investigations often require manipulation of large quantities of time-

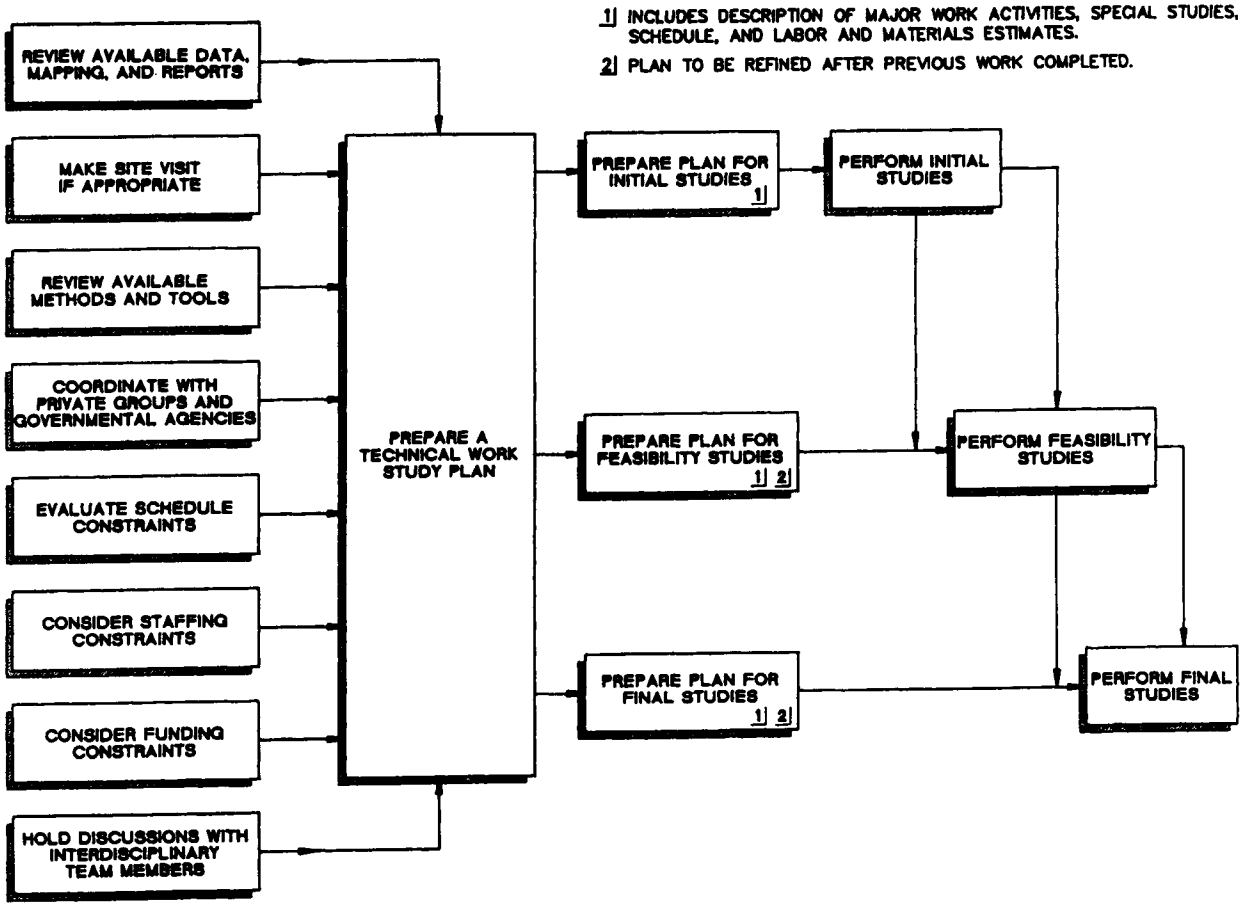


Figure 11.3.—Technical Study Work Plan.

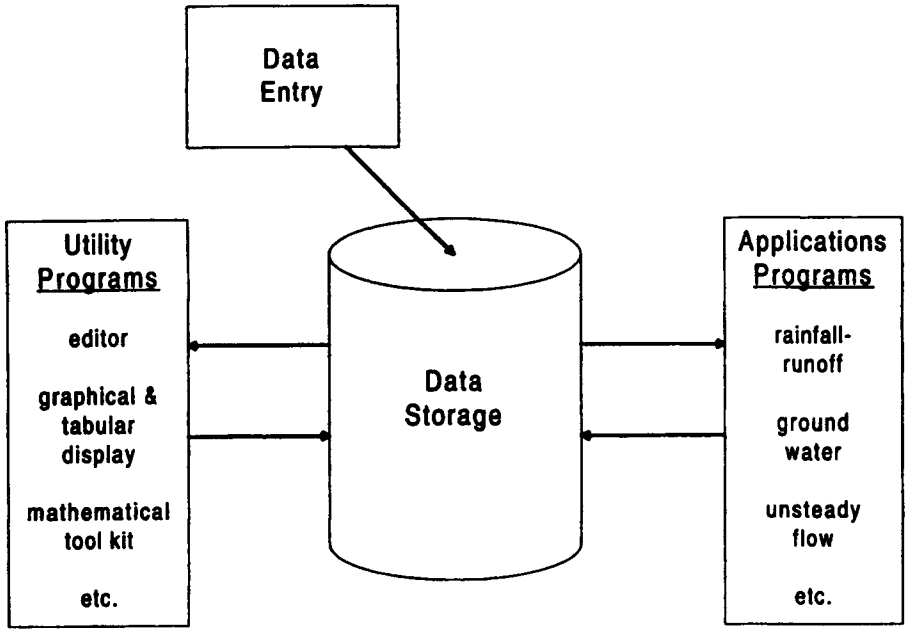


Figure 11.4.—Data Management for Hydrologic Studies.

series data, a data management system that deals with blocks of data can provide significant benefits. For example, individual blocks (records) of data might consist of daily-flow values for a year. By storing data in this manner, an entire year of daily flows can be handled as a single element and accessed with one "search" of the database. A reference by the U.S. Army Corps of Engineers (USACE, 1990) describes a nonproprietary data management system designed for efficient handling of time-series data for which a number of utility programs have been developed.

Data collection and processing are generally essential elements of a hydrologic investigation. Fortunately, much data are available on electronic media (e.g., compact disk), so that data entry can be performed relatively efficiently. For example, software can be used to retrieve data that is in the United States Geological Survey's WATSTORE format, in formats used by the National Weather Service, or in the Standard Hydrological Exchange Format used by various federal agencies (NWS, 1983). Once data have been entered into data storage, utility programs can be used to plot, tabulate, or develop statistical summaries of the data. Such initial analysis can greatly facilitate subsequent use of the data.

Fig. 11.4 shows applications programs communicating with the "data storage." This communication requires that such programs be developed or adapted with specific capabilities for this purpose (USACE, 1991). Alternatively, pre- and post-processors can be developed to enable such communication.

B. Geographic Information Systems (GIS)

A geographic information system (GIS) has been defined in various ways, usually related to its ability to process spatial data through capture, storage, retrieval, analysis, and display. A GIS, apart from computer mapping systems, does not simply display input-mapped data but creates new information through updating, query, and modeling capabilities. The most significant attribute of a GIS is its ability to relate pieces of spatial information with a relational database. This facilitates efficient handling of two generic classes of spatial data: cartographic data, describing the location and topography of map features, and attribute data, describing the characteristics of these features.

Hydrology is concerned with spatial data, namely the movement of water through space and time. Since a GIS is designed to manage and manipulate spatial data, such a system provides the tools to better assess and model the spatial relationships inherent in the hydrologic process. For example, a GIS may be

used to evaluate how land use changes may affect the movement of water through a watershed. A GIS enables the user to manipulate data and then view the results graphically and statistically all from one database system.

1. Emergence of GIS Computer mapping and analysis systems have been in existence since the early 1960s when grid-based software packages were developed. The 1970s brought larger and more complex systems such as the Dual Independent Map Encoded (DIME) file, developed by the U.S. Census Bureau, and SYMAP, developed at Harvard University. As more complex software packages were developed, the hardware necessary to run the statistical calculations and store the large data files could not keep pace with increasing processing requirements. However, within the last five to ten years, computer technology has improved and what once required a large main frame computer can now be stored in much smaller and affordable desk top workstations, personal computers, and disk or tape drives. With this capability came the ability to run sophisticated GIS packages such as ARC/INFO (Environmental System Research Institute (ERSI), Redlands, California) on relatively inexpensive hardware platforms. Smaller public agencies and private companies have taken advantage of the technology because affordability, technological limitations and disk space are no longer limiting factors.

2. Definition and Basic Components of a GIS There is a wide variety of computer software products on the market referred to as GIS, but a true GIS must contain all of the following major components (Marble, 1987):

a) A data input subsystem that collects and/or processes spatial data derived from existing maps, remote sensing, or other digital data files.
b) A data storage, retrieval, and update subsystem that organizes spatial data in a form permitting quick retrieval, as well as the ability to perform updates and modifications to the spatial data.
c) A data manipulation and analysis subsystem that performs a variety of tasks to create new data by changing the form of the data through user-defined aggregation, query, or overlay parameters.
d) A data reporting subsystem that is capable of displaying all or part of the original database as well as manipulating data and the output in tabular and map form.

What sets a GIS apart from computer mapping systems, such as Computer Aided Design (CAD) systems, is the ability to both automatically synthesize existing layers of geographic data and to update a database of spatial entities. This ability stems from the way information is stored in a GIS. A GIS is composed of two basic pieces of information: locational data represented by points, lines, polygons, or grid cells, and thematic attributes associated with these features. These two pieces of information are automatically linked in the GIS, and when either the mapped data or the tabular data are manipulated, both sets of data are updated and adjusted to maintain the relationship between them.

For example, a map of a hydrologic stream network is input into the GIS. Physical attributes used to describe and characterize a stream segment are assigned to the individual graphic element used to display the locational features of the stream segment. Certain attribute information, such as length, is automatically calculated while other attribute information, such as stream type, stream name and roughness coefficient, can be input as needed. The attribute data are stored in a file associated with or linked to the graphic (locational) information.

a. Data Structure Types. A GIS stores spatial data in one of two basic storage structures: raster and vector. Raster-based systems store data as a series of grid cells or in a picture element (pixel) format. A grid is essentially placed over the mapped features, and the dominant data type is assigned to an individual cell. Raster systems are more efficient in data storage and operation of analytical data manipulation; however, the resolution and accuracy of the data are limited to the size of the grid cell. The data resolution can be improved by reducing the cell size, but data volume subsequently increases. In

addition, graphic output may be unacceptable since the grid cells produce a stair–step effect, making it difficult to represent linear features. Since digital remote sensing data is in raster format, raster-based systems are used to process satellite data.

Vector-system structures store spatial data as a series of points, lines, and polygons that define point, linear, or aerial map features. Vector systems can portray any size of information and are limited by the resolution of the input data. The output displays from a vector-based system look like a standard map product.

The distinction between raster and vector systems is becoming increasingly fuzzy. Vector-based systems are integrating and adding raster-based capabilities, and vice–versa. Both raster and vector systems have distinct capabilities and the trend in software development is to integrate both approaches to take advantage of the capabilities inherent in both data structure types.

b. Data Layering. A key concept with GIS is data layering. Information is separated into "layers" based on a common thematic characteristic, such as soils, parcels, roads, well points, etc. Data layers are defined and differentiated depending upon the types of applications and the resolution of the data.

c. Coordinate Referencing. Data input to a GIS is internally referenced and externally registered to an earth-based coordinate system such as State Plane or Universal Transverse Mercator (UTM). Tic marks with known coordinate values are established, and all graphic features are related to these tic marks. This process provides a common reference between data layers, allowing the GIS to display separate data layers on top of each other. The accuracy of a data base can be highly dependent on how well map features are located in relation to a coordinate reference system. As a result, an important consideration with GIS database development is the quality of base maps and coordinate referencing.

d. Data Input. GIS data input and database development can be accomplished in a variety of ways including manual digitizing, scanning, data transfer, and data purchase. The most common form of data input is manual digitizing. Digitizing involves the use of a digitizing table and a hand-held cursor. The user traces the map features with the cursor, and the computer stores the information as coordinates. Digitized input can be labor-intensive, and the quality of the data is dependent on the digitizer's skill.

The use of scanning has increased with improved technology in recent years. The immediate product from a scanner is a pixel image that can be used as a graphic or picture background if needed. If data from the map are to be assigned attribute information and be used for GIS analysis, the scanned image must be vectorized (i.e., defined as points and lines in separate data layers). Although the capability of the computer to distinguish information from scanned images has greatly improved, a human operator is still required to help the computer make decisions.

Existing digital information can also be transferred into a GIS database from other computerized mapping systems (such as CAD) and between GIS databases. Data transfers save data input time but can be costly to acquire as well as time-consuming to convert into a compatible format. Commonly used data transfer formats include DXF, DLG, and ASCII. Because there are often fundamental differences between computerized mapping software programs, data transfers can result in the loss of certain types of data.

Digital data can also be purchased from a number of data sources. The United States Geological Survey (USGS) sells planimetric data (roads, streams, jurisdictional boundaries) in DLG format and elevation data as Digital Elevation Models (DEM). The U.S. Census TIGER files include road and address information as well as demographic and economic information. The U.S. Soil Conservation Service is in the process of preparing soils data in digital format. In addition, local state, county, and city agencies may have digital data, such as land use or land ownership, available for purchase. Issues related to using existing digital files include proprietary issues, cost, compatibility, and quality of data.

3. GIS Capabilities A GIS includes many map manipulation and analysis capabilities that can be used in hydrologic applications. The following list displays GIS capabilities.

a) Spatial query involves selecting particular features using Boolean queries and/or algebraic manipulations. For instance, soils with a high erosion potential can be highlighted by selecting soil categories with a soil erosion value greater than a certain criterion.
b) Polygon–overlay analysis involves overlaying two or more data layers to create a new layer of information. For example, this process may be used to create a coverage representing constraints to development of a pipeline by combining and rating information on slope, soils, and geology.
c) Buffering involves creating a new polygon around a map feature or features based upon a specified distance value. An engineer or planner may use the GIS buffering capabilities (along with overlay techniques) to investigate the type of land uses that occur within 500 feet of a stream channel, for example.
d) Surface analysis involves the capabilities to generate slope and aspect information from elevation data. A GIS uses an algorithm such as triangulation and user-specified parameters to generate slope or aspect coverages that can then be overlain on other data layers such as geological units to identify suitable locations for buried pipes.
e) A GIS includes advanced computer graphic capabilities to allow the display of data in a variety of formats and quality. Results can be displayed graphically on the computer screen or on hard copy maps generated by pen or electrostatic plotters. Information can be displayed at any user-specified size or scale limited only by the graphic hardware device.

4. GIS Applications in Hydrology GIS technology provides a rich set of tools for processing data for hydrology-related applications. The ability of the GIS to manage, manipulate, and display spatial data provides the GIS user with new and enhanced ways to investigate a problem that are not necessarily available or practical with CAD or traditional manual methods. Because a significant percentage of the data collected and evaluated in relation to water resource issues have a spatial component, the potential for applying GIS to hydrologic modeling, ground water monitoring, watershed management, water quality assessment, and other hydrology applications is considerable.

The following describes some general categories of GIS applications to hydrology issues. A more detailed discussion of the application of GIS with hydrologic modeling (specifically the integration of GIS with the SCS curve number model for runoff volume estimation) is provided since this is a straightforward and common application of GIS technology to hydrology issues. Undoubtedly, the number and types of applications will continue to grow in the future.

a. Hydrologic Modeling. Many hydrologic models have parameters defined in terms of land use, land cover, soils, slope, and other factors that generally vary across space. The SCS runoff curve number model (described in Chapter 3) is a widely used and accepted model that considers the combined effects of soil type, cover type, treatment, antecedent moisture conditions, and other hydrologic conditions to predict runoff volume. Several recent studies have used GIS to a variable degree to help store and manipulate soils, land cover, and other hydrologic input data (Stuebe and Johnston, 1990; Warwick and Haness, 1994). In some instances, the GIS was used simply for storing data on physical land characteristics and rainfall data, and an associated software program was used to retrieve information from the GIS to compute several runoff parameters. In other studies, many of the calculations and computations were performed using the GIS.

To illustrate a practical application of GIS, the following outline describes typical ways that a GIS can be used to generate and store data associated with the SCS curve number model.

1) Input soils polygons from SCS soil survey maps by scanning, manually digitizing, or purchasing in digital form. Each soils polygon is labeled with its SCS soils classification identification. Using the tabular data management capabilities of the GIS, each polygon is reclassified into one of the four hydrologic soils groups used with the SCS method. Boundaries between like hydrologic soils groups are eliminated to create a hydrologic soils data layer.

2) Input land cover information from existing land use maps or through interpretation of satellite data. Remote sensing has been used successfully to derive land cover estimates for the SCS curve number model in past studies (Ragan and Jackson, 1980). Land cover data are reclassified and dissolved, if necessary, into the cover classification used for the SCS method.
3) Overlay soils and land cover data layers to identify areas with unique combinations of hydrologic soil groups and land cover attribute data (hydrologic response units). Based upon the attribute data, runoff curve numbers are assigned to each hydrologic response unit polygon. The GIS is then used to efficiently and accurately determine the total acreage of each hydrologic response unit per watershed. Each individual hydrologic response unit is divided by the area of the total watershed to determine a spatially-weighted runoff curve number. The curve numbers are input into the runoff equation to estimate runoff. The computations can be programmed into the GIS to be calculated automatically.

The advantages of using the GIS to generate watershed data include gains in precision and speed of calculation over manual procedures (Berry and Sailor, 1987). The benefits of using a computerized system are most evident with repeated iterations of calculations. Repeated simulations of a variety of proposed levels of development can be made rapidly, and the effects on runoff can be assessed and better understood. The "what if" questions a GIS can effectively answer provides valuable insights into the spatial ramifications of planned activities.

b. Other Water Resource Applications. The spatial processes related to water quality are similar to hydrologic (runoff) processes. Typically, the same input variables used for hydrologic modeling (land use, land cover, soils, topography) are also used as input variables for water quality and soil loss assessment and modeling. Several studies have used a GIS to implement the Universal Soil Loss Equation (USLE) or similar model types to calculate potential soil loss (Gilliland and Baxter–Potter, 1987; Evans and Miller, 1988; DelRegno and Atkinson, 1988). GIS has useful applications with respect to water quality modeling and nutrient loading assessment by helping to incorporate spatial factors such as distance to streams into the modeling process (Levine and Jones, 1990).

For less technical applications, GIS has been successfully applied to water resource management and planning programs that often deal with the review and manipulation of mapped resource data for large geographic areas (Hamlett et al., 1991; Baker and Panciera, 1990). The data storage and manipulation capabilities of the GIS allow users to view an entire study area at one time or to focus in on a particular area of interest. The ability to view and combine separate data layers to create new information is particularly effective for planning applications.

Remote sensing can provide information for large geographic areas and is directly compatible with raster-based GIS. The combined capabilities of remote sensing and GIS have been utilized to develop watershed-level information on evapotranspiration, canopy cover, land use, and vegetation.

Most of the applications described above deal with polygon- (or area-) related information. GIS is also an effective tool for handling line or network information sets such as water distribution pipelines or stormwater drainage networks. GIS technology has been used to prioritize, schedule, and coordinate infrastructure maintenance programs. Urban runoff pollution studies have used GIS to track the locations and characteristics of storm drain outfalls.

C. Conclusion

Existing data management technology has the potential to dramatically alter the way hydrologic studies are performed and to contribute to new levels of understanding of hydrologic processes (Lanfear, 1989). As the technology continues to improve, the ability to efficiently manage the tremendous quantities of spatial data associated with hydrologic modeling and analysis will likely allow for a more complete consideration of the spatial relationships inherent in hydrologic processes.

IV. CALIBRATION AND VERIFICATION OF HYDROLOGIC MODELS

A. What is Calibration?

Earlier chapters of this handbook describe various models for analysis of hydrologic systems. All of these models "interrelate, in some time reference, an input, cause or stimulus, of matter, energy, or information to an output, effect, or response, of information, energy, or matter . . ." (Dooge, 1973). Some of the models described are causal—the interrelationship of input to output is based on functional representation of the laws of thermodynamics or conservation of mass, momentum, or energy. The unsteady streamflow routing models of Chapter 6 are causal. They solve equations that describe motion of fluid in an open channel. Other models are empirical—they represent only the functional relationship of observed output to observed input. The direct correlation model described in Chapter 8 for estimating runoff volume is an example of such an empirical model. The Chapter 8 model does not consider the process by which rainfall is converted to runoff. Instead, it uses historical data to establish a relationship with which an engineer can predict runoff, given rainfall estimates.

To use a selected model of either type, the analyst must estimate parameters of the functions that relate output to input. A parameter represents a property or characteristic of the system under specified conditions (McCuen and Snyder, 1986). For example, in the ground water models presented in Chapter 5, hydraulic conductivity is a parameter that describes a property of the aquifer. Model parameters may be classified as physically-based parameters or as calibration parameters. Physically-based parameters are those that can be observed or estimated directly from measurements of the system characteristics. Calibration parameters, on the other hand, are lumped single-valued parameters that cannot be quantified from direct measurements. They represent integrated, spatially- and temporally-averaged conceptual approximations to system components.

How each calibration parameter is estimated depends on the data available. If input and simultaneously-observed output data are available, the calibration parameters may be found by comparing results with alternative parameter values to find those that optimize the fit of predicted output to observed output. For example, the parameters of a unit hydrograph model may be found by computing runoff hydrographs with alternative calibration parameter values. These computed hydrographs are compared with the observed hydrograph. The "best" parameters are those that yield the "best" fit. This optimization procedure is described in further detail in the following subsection.

Unfortunately, as Loague and Freeze (1985) point out, "when it comes to models and data sets, there is a surprisingly small intersecting set." That is, the input and output data necessary to search for the calibration parameters often are not available. This is referred to herein as an ungaged case. Data may be missing, sparse, or unreliable. Furthermore, for many studies, estimates of hydrologic system output are required for the forecasted future and for with-project conditions. Data are never available to permit parameter estimation via optimization for these conditions. When data are available at other sites in the homogeneous region where the estimate is required, a model can be fitted for use at ungaged sites in the region. The section titled Estimating Model Parameters in the Ungaged Case presents alternatives for parameter estimation for such cases.

B. Calibrating a Model with Process Input and Output Data Available

If observed system input and simultaneously-observed output data are available, the calibration parameters can be estimated via model parameter optimization. To accomplish this, the analyst must select 1) the criterion for measuring fit and 2) the procedure to identify the best (optimal) parameter estimates, given the criterion selected.

1. Criteria for Measuring Fit The assessment of goodness of fit may be either subjective, such as is possible with a graphical comparison, or objective, such as is possible with statistical indices. The purpose of such assessment is to evaluate the accuracy of a model. Factors that cause inaccuracy in model

predictions may introduce either systematic or nonsystematic (i.e., random) errors. Bias is the term used to indicate systematic errors; precision is the term used to indicate random errors. Accuracy is the cumulative effect of bias and precision.

A visual comparison of observed values to values predicted with alternative parameters provides a quick assessment of model fit. Certain errors, such as those in peak flow rates or volumes, are evident immediately from such a comparison. Further, graphical comparison often provides the analyst with a better understanding of model capabilities and sensitivity to parameters. Figs. 11.5a to d show four plots that illustrate combinations of the effects of bias and imprecision, with the 45° line indicating equality of predicted and measured magnitudes. Both Fig. 11.5b and 11.5d show a tendency to overpredict; thus, they show a positive bias. Neither Fig. 11.5a nor 11.5c show a significant overprediction or underprediction. Thus, they are unbiased since there is no systematic error evident in the plots.

While Fig. 11.5c may indicate that the prediction model is not biased, there still is sufficient scatter that suggests the model is not accurate, i.e., the model does not yield precise estimates. Thus, the scatter of points in Fig. 11.5c indicates a model that provides unbiased but imprecise estimates. The scatter of Fig. 11.5a indicates a prediction model that provides precise, unbiased estimates; thus, it is an accurate model. The scatter of Fig. 11.5b suggests that the model provides precise but biased estimates because there is minimal scatter about the central tendency (dashed line). The model for Fig. 11.5d provides biased, imprecise estimates and is thus highly inaccurate.

In addition to the overall biases evident in Figs. 11.5b and 11.5d, models may provide local biases. That is, a model may provide accurate estimates for some range of predicted values but inaccurate values in other ranges. For example, in Fig. 11.5e, the model provides reasonable estimates for low values but shows an increasing bias for the large values. In Fig. 11.5f, the bias is negative for high values of the magnitude and positive for low values. The model of Fig. 11.5d always overpredicts because the predicted value is greater than the measured value.

Other types of data may require other forms of graphical analysis; however, in these cases, bias, precision, and accuracy should still be assessed. James and Burges (1982) suggest four types of plots for comparing model results with recorded data:

a) A continuous time-series plot of predicted flow versus observed flow. Fig. 11.6 is an example of such a plot. The model provides biased estimates of the flood peaks; all of the measured peaks are underpredicted by the model. However, the model gives unbiased, precise estimates of the recession rate of the hydrograph. The baseflow also appears to be unbiased. Based on these subjective assessments of the goodness of fit, model coefficients related to peak discharges need to be adjusted to eliminate the bias. Recession and baseflow coefficients appear to be accurate and do not require adjustment.
b) A continuous time-series plot of the difference between observed and predicted flow. This is similar to the first plot, but the significance of differences is more obvious. With this plot, local biases can be detected. This can lead to improved parameter estimation.
c) A plot of cumulative sum of departures from the observed mean for the predicted series versus cumulative sum of departures of the observed series. This will show the cumulative impact of errors and may lead to improved parameter estimates. For example, if the infiltration parameter of a rainfall–runoff model is incorrectly estimated, the error in the predicted total volume will accumulate. This can be seen with a plot of cumulative departures.
d) A scattergram of predicted versus measured values, which was discussed previously.

These plots can be prepared to compare results from models of other hydrologic systems. Although assessments made with graphical comparisons are subjective, they do provide information regarding goodness–of–fit that may not otherwise be apparent with statistical measures.

A variety of numerical measures have been proposed to objectively quantify the fit of predicted to measured values. These indices are typically used for one of two functions: single-valued variables or

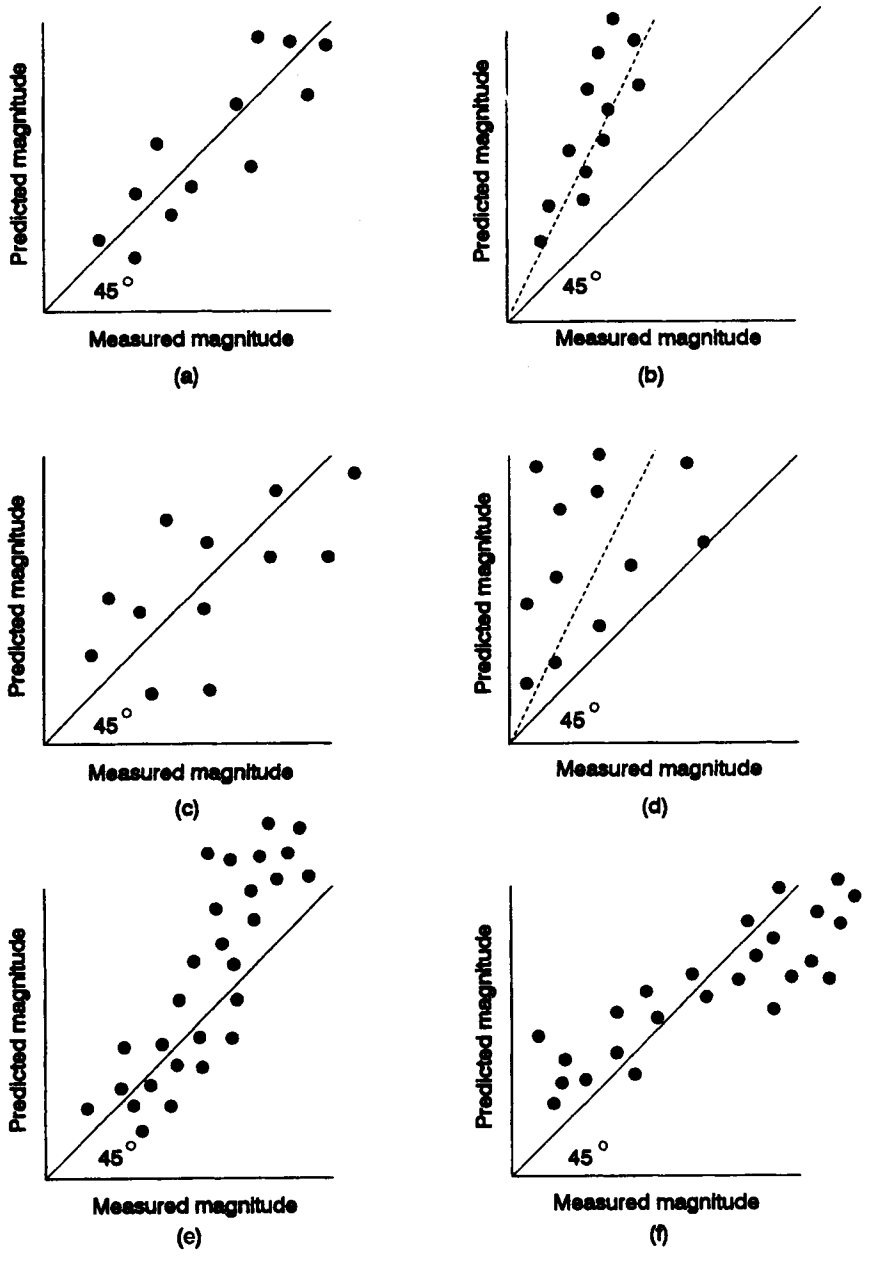


Figure 11.5.—Scattergrams of Measured vs. Predicted Magnitudes.

multi-ordinate functions. The error in the peak discharge of a runoff hydrograph is an example of the single-valued variable. The sum of squares of the differences between the predicted and measured hydrograph ordinates is an example of the multi-ordinate function. In addition to the type of function, the indices can be presented in either dimensioned or dimensionless form. For example, the error in the peak discharge relative to the measured peak discharge would be dimensionless. When comparing different model outputs, such as peak discharges and volumes of runoff, especially for cases of different storm magnitudes, dimensionless values make for more rational assessments. Dimensioned values are preferred when assessing the engineering effects of an error. For example, if the design engineer knows the expected error in a predicted peak discharge is X m^3/s, then the magnitude of the error may influence the design.

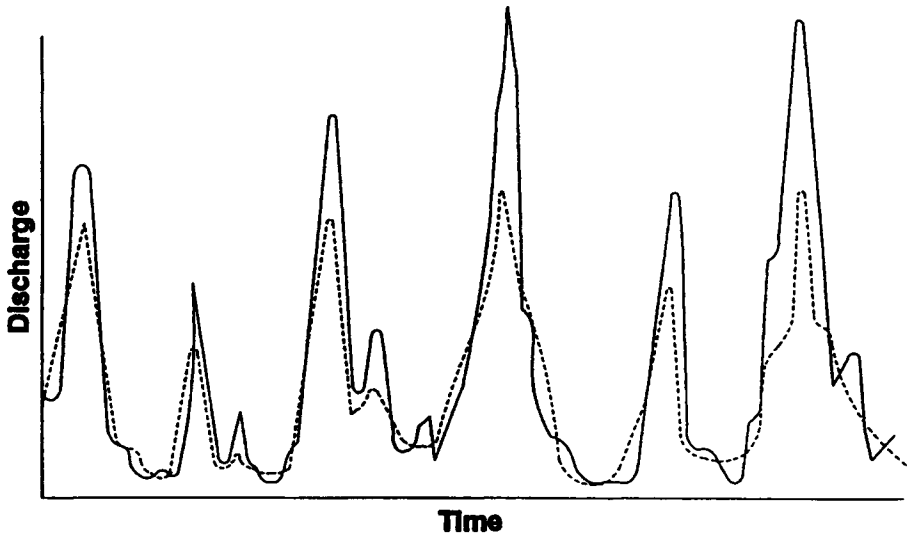


Figure 11.6.—Graphical Comparison of Observed and Predicted Time Series.

The goodness-of-fit indices may be used to measure either bias or accuracy. General forms of indices are given in Table 11.1; specific examples are given in Table 11.2 (Green and Stephenson, 1986. Reprinted with permission from IAHS Press).

With frequency or probability models, the criteria for comparison are slightly different, as direct comparison of time series is not applicable. Instead, statistics or probability estimates of the observed series are compared with those predicted with a chosen model. For example, the method of moments identifies the optimal parameters of the normal probability model by seeking parameters that will yield predicted values of the mean and standard deviation that are equal to the mean and standard deviation of the observed data. Similarly, the method of maximum likelihood seeks parameters that will maximize the probability of predicting the observed values, given the form of the model. However, objective criteria for assessing the goodness-of-fit of data to a computed frequency curve are rarely used; generally, a visual, subjective assessment of goodness-of-fit is the method of choice.

2. ***Procedures for Identifying Optimal Parameter Estimates*** The procedure employed to identify the best optimal parameter estimates, given the criterion selected, depends on the mathematical characteristics of the model and of the criterion. Himmelblau (1972) identifies the following general classes of optimization methods appropriate for model-parameter estimation:

a) *Graphical methods* in which a plot of the criterion as a function of the parameters is prepared, and the optimal parameters are identified by inspection of the plot. These methods are limited to models with one or two variables. Helsel and Hirsh (1992) provide a comprehensive discussion of graphical analysis.
b) *Analytical methods* that make use of classical calculus techniques. The method of ordinary least squares used to estimate parameters of a linear model with the minimum sum of squares criterion is an example. In that case, the optimization method finds parameters that cause the derivative of the criterion function with respect to the parameters to vanish. Haan (1977) provides background information and applications on analytical methods.
c) *Case study methods* in which the criterion is evaluated for a number of alternative parameters. From this set, the "best" is selected.
d) *Numerical methods* that use past information to generate better solutions to the optimization problem by means of iterative procedures. Numerical methods can be used to identify optimal parameters in cases where other methods are unusable. Numerical optimization methods are described in

TABLE 11.1. Goodness-of-Fit Assessment Criteria—General Forms.

Type of function (1)	Form of function (2)	Function (3)	Measure of (4)
Single-valued	Dimensioned	$e = (\hat{X} - X)$	Bias
	Dimensionless	$e = \frac{(\hat{X} - X)}{X}$	Relative bias
Multi-ordinate	Dimensioned	$\bar{e} = \frac{1}{n}\sum_{i=1}^{n}(\hat{X}_i - X_i)$	Bias
	Dimensionless	$\bar{e}_r = \frac{1}{n\bar{X}}\sum_{i=1}^{n}(\hat{X}_i - X_i)$	Relative bias
	Dimensioned	$S_e = \left[\frac{1}{n-p}\sum_{i=1}^{n}(\hat{X}_i - X_i)^2\right]^{0.5}$	Accuracy
	Dimensionless	$\frac{S_e}{S_x} = \frac{\left[\sum_{i=1}^{n}(\hat{X}_i - X_i)^2\right]^{0.5}}{\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^{0.5}}$	Relative accuracy

n number of pairs of ordinates compared in a single event
p number of predictor variables
X observed value
$\hat{X}$ estimated value
$\bar{X}$ mean of the observed values

a variety of operations research texts. McCuen (1992) applies a numerical algorithm for the solution of hydrologic analyses.

Table 11.3 and Fig. 11.7 illustrate the search for the optimal parameter value for a simple rainfall runoff model of a 295 km^2 watershed. The model used in this example is a tank model. The tank (reservoir) represents the lumped effects of all storage in the watershed. Tank inflow is the product of areal–average rainfall (column 2) and area, and simulated runoff, $q_s(t)$, is the tank outflow. The simulated outflow is a linear function of average storage, $S(t)$, as defined by $q_s(t) = kS(t)$. In this model, k is a calibration parameter; it must be found by fitting the model to available data. To do so, a trial value, $k = 0.00$/h, was selected, and the corresponding runoff hydrograph was computed with the model. That hydrograph is tabulated in column 4. The goodness-of-fit with this parameter estimate was quantified by computing squared residuals (column 5) and summing these to find G (see Table 11.2). This process was repeated for other values of k, and G was plotted vs. k in Fig. 11.7. When G is minimized, the fit is "optimal," and the corresponding value of the parameter k could be used with the model to predict output from other input. By inspecting the plot, one could conclude that $k = 0.03$/hr is the optimal value; additional trials would "home in" on $k = 0.033$/hr.

C. Estimating Model Parameters in the Ungaged Case

To estimate hydrologic system output in cases in which observed input and output data are not available, the analyst can do the following: 1) use a model that includes only parameters that can be observed or inferred from measurements, or 2) extrapolate calibration parameters from those found for gaged systems. In practice, some combination of these solutions typically is employed, because most models include both physically-based and calibration parameters.

TABLE 11.2. Goodness-of-Fit Assessment Criteria—Examples (Green and Stephenson, 1986, Reprinted with permission from IAHS Press).

No. (1)	Criterion (2)	Equation (3)	Remarks (4)	Reference (5)
1	Sum of squared residuals	$G = \sum_{i=1}^{n} \left[q_o(t) - q_s(t) \right]_i^2$	Most common	Diskin & Simon (1977)
2	Sum of m-powered residuals	$G = \sum_{i=1}^{n} \left[\left(q_o(t) \right)^b - \left(q_s(t) \right)^b \right]_i^c$	Changes vertical scaling	Johnston & Pilgrim (1976)
3	Sum of absolute errors	$G = \sum_{i=1}^{n} \left\| q_o(t) - q_s(t) \right\|_i$	Suited to optimization	Stephenson (1979)
4	Model efficiency	$R^2 = \frac{F_o^2 - F^2}{F_o^2}$ with $F^2 = \sum_{i=1}^{n} \left[q_o(t) - q_s(t) \right]_i^2$ and $F_o^2 = \sum_{i=1}^{n} \left[q_o(t) - \bar{q} \right]_i^2$	Dimensionless	Nash & Sutcliffe (1970)
5	Normalized objective function	$P = \frac{1}{\bar{q}} \left(\frac{F^2}{n} \right)^{1/2}$	Coefficient of variance	Ibbitt & O'Donnell (1971)
6	Root mean square error	$RMSE = \left(\frac{1}{n} \sum_{i=1}^{n} \left(q_o(t) - q_s(t) \right)_i^2 \right)^{1/2}$	$P = \frac{RMSE}{\bar{q}}$	Patry & Marino (1983)
7	Automatic parameter optimization	$G = \sum_{i=1}^{n} \left[\frac{q_o(t) - q_s(t)}{\left(q_o(t) \right)^b} \right]_i^c$	b, c are control parameters	Wood (1974)
8	Reduced error estimate	$REE = \left[\frac{\sum_{i=1}^{n} \left(q_o(t) - q_s(t) \right)_i^2}{\sum_{i=1}^{n} \left(q_o(t) - \bar{q} \right)_i^2} \right]^{1/2}$	Biased to larger flows	Manley (1978)
9	Proportional error of estimate	$PEE = \left[\sum_{i=1}^{n} \left(\frac{q_o(t) - q_s(t)}{\left(q_o(t) \right)} \right)_i^2 \right]^{1/2}$	Equal weight to equal proportional errors	Manley (1978)

10	Standard error of estimate	$SEE = \left(\sum_{i=1}^{n} \frac{\left(q_o(t) - q_s(t)\right)_i^2}{(n-2)} \right)^{1/2}$	Dimensional and independent of number of points	Jewell et al. (1978)
11	Coefficient of persistence	$CP = \sum_{i=1}^{n} \left(\frac{a^2}{F^2} \right)_k$	Dimensional	Wallis & Todini (1975)
12	Percent error in peak	$PEP = \frac{q_{ps} - q_{po}}{q_{po}} \text{ X } 100$		
13	Percent error in volume	$PEV = \frac{V_s - V_o}{V_o}$		
14	Percent error in mean	$PEM = \frac{\bar{q}_s - \bar{q}_o}{q_o}$		
15	Total overall sum of squared residuals	$TSSR = \sum_{j=1}^{n} \left\{ \sum_{i=1}^{n} \left[q_o(t) - q_s(t) \right]_i^2 \right\}_j$		
16	Total overall sum of absolute residuals	$TSAR = \sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} \left\| q_o(t) - q_s(t) \right\|_i \right\}_j$		
17	Sum of absolute areas of divergence	$A = \sum_{i=1}^{n} \left\| \frac{(residual)_i + (residual)_{i+1}}{2} \Delta t \right\|_i$		
18	Total sum of absolute areas of divergence	$TSAA = \sum_{j=1}^{m} A_j$		
19	Variance	$S^2 = \frac{1}{n} \sum_{i=1}^{n} \left[q_o(t) - q_s(t) \right]_i^2$		
20	Mean deviation	$MD = \frac{1}{n} \sum_{i=1}^{n} \left[q_o(t) - q_s(t) \right]_i$		
21	Absolute area/ordinate divergence	$\frac{A}{n}$		

a area of an individual segment of deviation
k number of runs of successive over- or under-predictions
m number of events
n number of pairs of ordinates compared in a single event
q_o (t) observed flow rate at time t
q_s (t) simulated flow rate at time t
$\bar{q}$, $\bar{q}_o$ mean of observed flow rate
$\bar{q}_s$ mean of simulated flow rate
q_{pc} observed peak flow rate
q_{ps} simulated peak flow rate
V_o observed volume
V_s simulated volume

TABLE 11.3. Rainfall and Discharge Data For Calibration Example.

Time (h) (1)	Avg. rainfall for period (mm) (2)	$q_o(t)$ (m^3/s) (3)	$q_o(t)$ with k = 0.0 (m^3/s) (4)	Squared residual (5)
0	1.57	2.22	0	4.93
3	1.92	5.8	0	33.64
6	2.39	11.9	0	141.61
9	0	15.2	0	231.04
12	0	15.9	0	252.81
15	0	12.7	0	161.29
18	0.14	10.5	0	110.25
21	1.64	10.2	0	104.04
24	0	10.9	0	118.81
27	0	12.3	0	151.29
30	0	11.2	0	125.44
33	0	9.6	0	92.16
36	0	9.0	0	81.00
39	0	8.4	0	70.56
42	0	8.0	0	64.00
45	0	7.4	0	54.76
48	0	6.7	0	44.89
51	0	6.1	0	37.21
54	0	5.5	0	30.25
57	0	5.1	0	26.01
60	0	4.6	0	21.16
63	0	4.1	0	16.81
66	0	3.6	0	12.96
69	0	3.1	0	9.61
72	0	2.6	0	6.76
75	0	2.1	0	4.41
78	0	1.7	0	2.89
81	0	1.3	0	1.69
84	0	0.9	0	0.81
87	0	0.6	0	0.36
90	0	0.3	0	0.09
			Sum, G	2013.54

With the first solution, all model parameters can be estimated directly from measurable characteristics. For example, the parameters of the Muskingum–Cunge streamflow routing model (Cunge, 1969; Dooge et al., 1982; Ponce, 1989) are channel geometry, reach length, roughness coefficient, and slope. These parameters may be estimated with topographic maps, field surveys, photographs, and site visits. Therefore, this model may be used for analysis of an ungaged system.

With the second solution, parameters are estimated indirectly through extrapolation of gaged-system results. This extrapolation is accomplished by developing equations that predict the calibration parameters for the gaged systems as a function of measurable system characteristics. The assumption is that the resulting predictive equations apply for systems other than those from which data are drawn for development of the equations but have the same characteristics as the gaged sites. For example, calibration-parameter predictive equations for a rainfall–runoff model of an ungaged watershed might be developed as follows:

1) Collect rainfall and discharge data for gaged watersheds in the region. The watersheds selected should have hydrological characteristics similar to the ungaged watershed of interest. For example, the gaged and ungaged watersheds should have similar geomorphological and topographical characteristics. They should have similar land use, vegetative cover, and agricultural practices. The

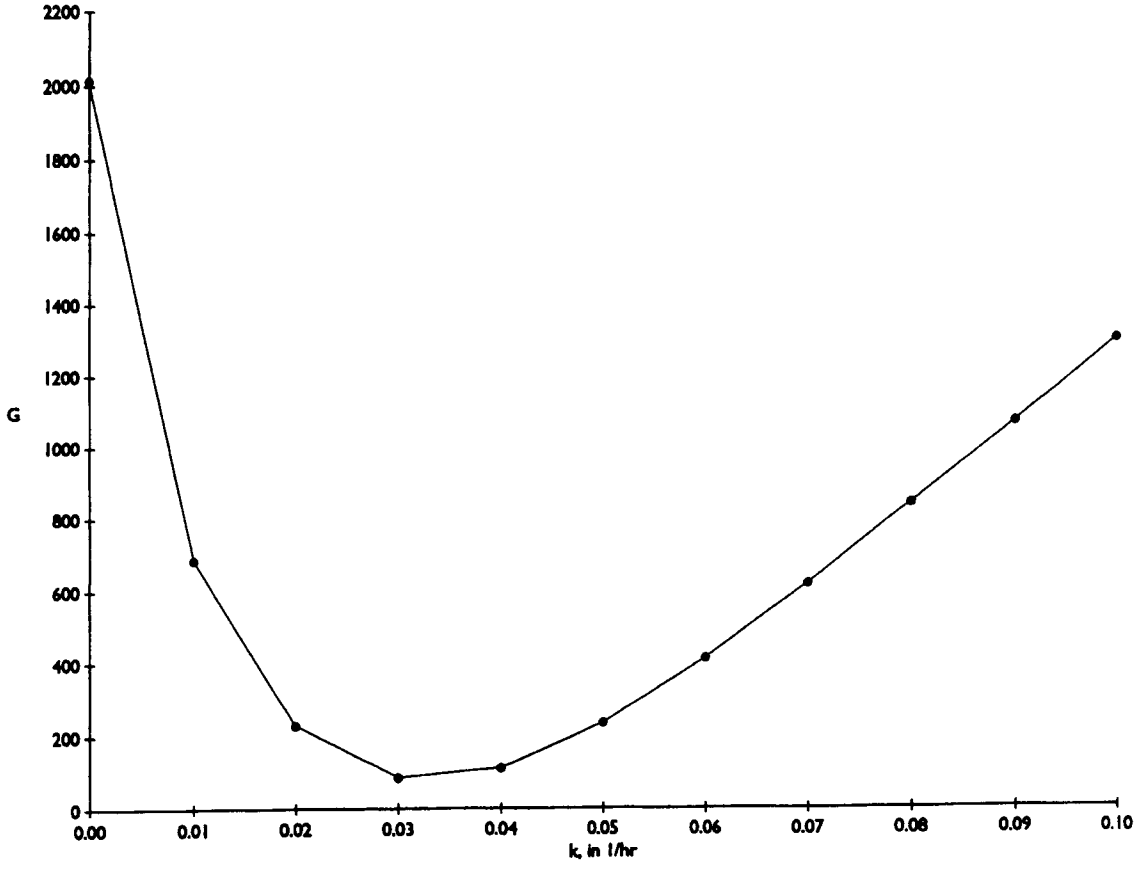


Figure 11.7.—Sum of Squared Residuals vs. Tank-Model Parameter.

watersheds should be of similar size. Rainfall distribution and magnitude and factors affecting rainfall losses should also be similar. If possible, data should be collected for several flood events. These rainfall and discharge data should represent, if possible, events consistent with the intended use of the model of the ungaged watershed. If the rainfall–runoff model will be used to predict runoff from large design storms, data from large historical storms should be used to estimate the calibration parameters.

2) For each gaged watershed individually, use the data to estimate the calibration parameters for the selected rainfall–runoff model. Procedures for doing so are described earlier.
3) Select and measure or estimate physiographic characteristics of the gaged watershed to which the rainfall–runoff model parameters may be related. Table 11.4 lists candidate watershed characteristics. Some of these characteristics, such as the watershed area, are directly measured. Others, such as the Horton ratios, are computed from measured characteristics.
4) Develop predictive equations that relate the calibration parameters found in step 2 (the dependent variables) with characteristics measured or estimated in step 3 (the independent variables). In a simple case, the results of steps 2 and 3 may be plotted with a rainfall–runoff model parameter on the ordinate and a watershed characteristic selected in step 3 on the abscissa. Each point of the plot represents the value of the parameter and the selected characteristic for one gaged watershed. With such a plot, a relationship can be visually fitted and sketched on the plot. Regression analysis is an alternative to the subjective graphical approach to defining a predictive relationship. Regression procedures determine analytically the optimal predictive equation. Details of regression analysis are presented in hydrologic statistics texts, including those by Haan (1977) and McCuen and Snyder (1986). Table 11.5 shows example equations derived for unit-hydrograph parameters.

TABLE 11.4. Watershed Characteristics for Calibration-Parameter Predictive Equations.

Total watershed area
Area below lowest detention storage
Stream length
Stream length to watershed centroid
Watershed shape (e.g., circularity ratio)
Average watershed slope
Average conveyance slope
Conveyance slope measured at 10% and 85% of stream length (from mouth)
Height differential
Elevation of watershed centroid
Average of elevation of points at 10% and 85% of stream length
Permeability of soil profile
Soil–moisture capacity average over soil profile
Hydrologic soil group
Population density
Street density
Impervious area
Directly-connected impervious area
Area drained by storm sewer system
Land use
Detention storage
Rainfall depth for specified frequency, duration
Rainfall intensity for specified frequency, duration
Horton's ratios (Horton, 1945)
Drainage density (Smart, 1972)
Length of overland flow (Smart, 1972)

To apply a parameter-predictive equation for an ungaged system, the independent variables are measured or estimated for the ungaged system. Solution of the equation with these values yields the desired calibration parameter. This parameter is used then to predict output from the ungaged system.

In some hydrologic engineering studies, the goal is to define frequency relationships. Chapter 2 describes procedures for precipitation frequency studies, and Chapter 8 summarizes procedures and describes statistical models for discharge frequency analysis. All of the models described are empirical.

TABLE 11.5. Example Calibration-Parameter Prediction Equations for Unit Hydrograph Model.

Equation	Reference
$C_t = 7.81/I^{0.78}$	WME (1969)
$C_p = 0.89\ C_t^{0.46}$	WME (1969)
$R = c\ T_c$	Russell et al. (1979)
$T_c = 8.29\ (1.00 + I)^{-1.28}\ (A/S)^{0.28}$	USACE (1982)

Note: In the above equations, C_t = calibration coefficient for Snyder's unit hydrograph (UH); C_p = calibration coefficient for Snyder's UH; T_c = time of concentration (h); R = Clark's instantaneous UH storage coefficient (h); I = impervious area (%); L = length of channel/ditch from headwater to outlet (mi); S = average watershed slope (ft/ft); c = coefficient, for forested watershed = 8–12, for rural watersheds = 1.5–2.8, and for developed watershed = 1.1–2.1; A = watershed area (mi^2).

Observed data are necessary for calibration. Consequently, these statistical models cannot be applied directly to an ungaged watershed. Options available to the analyst requiring frequency estimates for an ungaged stream include the following: 1) develop frequency-distribution parameter-predictive equations, or 2) develop distribution-quantile-predictive equations.

Procedures used by the Corps of Engineers illustrate the first option. The Corps and other federal agencies use the log–Pearson type III statistical distribution (model) for annual maximum discharge-frequency studies. This model has three parameters. Following federal guidelines (USWRC, 1982), these parameters are estimated from the mean, standard deviation, and skew coefficient of the logarithms of observed peak discharges. In the absence of flow data, regional frequency analysis procedures may be applied to develop parameter-predictive equations. As with the equations for rainfall–runoff model parameters, these equations relate model parameters to watershed characteristics. For example, for the Shellpot Creek watershed, Delaware, the following predictive equation was developed (USACE, 1982):

$$S = 0.311 - 0.05 \log A \tag{11.1}$$

in which S is the standard deviation of logarithms and A is the drainage area, in sq. mi. With similar equations, other parameters can be estimated. To apply a distribution parameter-predictive equation for an ungaged watershed, the independent variables in the equation are measured or estimated for the ungaged watershed. Solution of the equation with these values yields the desired statistical-distribution parameter. The frequency curve is computed with the model.

With the second option, the frequency distribution quantiles for an ungaged watershed are defined directly with predictive equations. Such a predictive equation is developed by defining the frequency distributions for streams with gaged data, identifying from the distributions specified quantiles, and using regression analysis procedures to derive a predictive equation. For example, for the Red Lion Creek watershed, Delaware, the following quantile predictive equation was developed (USACE, 1982):

$$Q_{100} = 1040\, A^{0.91} \tag{11.2}$$

in which Q_{100} is 100-year (0.01 exceedance probability) discharge and A is drainage area in sq. mi. To use this or a similar predictive equation for an ungaged watershed, the independent variable(s) in the equation are measured or estimated for the ungaged watershed. Solution of the equation yields the desired quantile. The USGS publishes regional regression equations for virtually the entire United States.

D. Validating Estimated Model Parameters

The final stage in preparing for model application is validation of estimated model parameters. If input and output data are available, the predictive ability of the model should be evaluated with data other than those used in calibration-parameter estimation. Criteria for goodness-of-fit in calibration can be employed for validation. If the model predicts well with these test data, the model and parameters can be accepted conditionally and used to predict hydrological system responses to design inputs. However, the validation must be an on-going activity, as the model and parameters must be reassessed as system conditions change and as additional data are available for calibration. For example, if the land use changes, the analyst must recalibrate the parameters (or perhaps, the model functional relationships) to reflect the modified conditions. The continuing process of model assessment with new data indicates if such modification is necessary.

Validation is more difficult in the ungaged case, as no data are available for comparison. Consequently, the analyst must be more creative in searching for confirmation of output estimates. One possible method of validation in the case of a rainfall–runoff model is illustrated in Fig. 11.8. This procedure was used in El Dorado County, California, to validate a rainfall–runoff model intended to estimate design flows of

CONTENTS

CONVERSTION TO SI UNITS

To Convert from	To	Multiply By
acre (ac)	hectare (ha)	0.4047
acres	square kilometer (km^2)	0.004047
acres	square meters (m^2)	4046.8726
acre-feet (ac-ft)	cubic meters (m^3)	1233.489
acre-feet	hectare-meter (ha-m)	0.123349
cubic foot (ft^3)	cubic meter (m^3)	0.02832
cubic feet per second (cfs)	cubic meters per second (m^3/s)	0.02832
cubic feet per second	cubic meters per minute (m^3/min)	1.69901
cubic yard	cubic meter (m^3)	0.7645
degree Fahrenheit (°F)	degree Celsius (°C) = (°F − 32) 1.8	
degree Kelven (K)	K − 273 + °C	
feet per second (ft/s)	meters per second (m/s)	0.3048
foot (ft)	meter (m)	0.3048
gallon, U.S. (gal)	cubic meters (m^3)	0.0038
gallon, U.S.	liter (l)	3.79
gallons, U.S. per minute (gpm)	cubic meters per second (m^3/s)	6.309×10^{-5}
horsepower (HP)	kilowatt (k)	0.746
inch (in)	millimeter (mm)	25.40
mile (mi)	kilometer (km)	1.609344
miles per hour (mi/hr)	kilometers per hour (km/hr)	1.609344
million gallons per day (mgd)	cubic meters per second (m^3/s)	0.0438
pound force (lbf)	newton (N)	4.45
pound force per inch (psi)	kilopascal (kpa)	6.89
pound mass (lbm)	kilogram (kg)	0.4536
short ton	kilogram (kg)	907.2
square foot (sq ft) (ft^2)	square meter (m^2)	0.0929
square inch (sq in) (in^2)	square millimeter (mm^2)	645.
square mile (sqm)	square kilometer (km^2)	2.590
square yard (sq yd) (yd^2)	square meter (m^2)	0.8361
yard (yd)	meter (m)	0.9144

LIST OF FIGURES

LIST OF TABLES

INDEX

C

E

F

G

W